

65	71	2.8	188914	9	AC019118	Homo sapi
66	71	2.8	189605	2	AC130472	Homo sapi
67	71	2.8	197709	2	AC073641	Homo sapi
68	71	2.8	198388	2	AC073426	Homo sapi
69	71	2.8	198398	2	AC022219	Homo sapi
70	71	2.8	198748	2	AC016719	Homo sapi
71	71	2.8	200000	2	AP000494	Homo sapi
72	71	2.8	200885	2	AC012460	Homo sapi
73	71	2.8	209382	2	AC005071	Homo sapi
74	71	2.8	211945	2	AC006238	Homo sapi
75	71	2.8	215004	2	AC068146	Homo sapi
76	71	2.8	285886	2	AL391377	Homo sapi
77	71	2.8	303189	2	AC108728	Homo sapi
78	71	2.8	340000	2	AP001712	Homo sapi
79	71	2.8	104972	2	AC025176	Homo sapi
80	70	2.7	119185	2	AL157699	Homo sapi
81	70	2.7	135752	2	AC073465	Homo sapi
82	70	2.7	146802	2	AC025186	Homo sapi
83	70	2.7	172283	2	AC040942	Homo sapi
84	70	2.7	180948	2	AC095516	Homo sapi
85	70	2.7	181113	2	AC114399	Homo sapi
86	70	2.7	183584	2	AC012492	Homo sapi
87	70	2.7	208561	2	AC023206	Homo sapi
88	70	2.7	211063	2	AC069254	Homo sapi
89	70	2.7	30257	2	AL603962	Homo sapi
90	69	2.7	42213	2	AC010527	Homo sapi
91	69	2.7	44735	2	AC092315	Homo sapi
92	69	2.7	48623	2	AC130469	Homo sapi
93	69	2.7	49372	2	AC109812	Homo sapi
94	69	2.7	60310	2	AL391647	Homo sapi
95	69	2.7	63753	2	AC090089	Homo sapi
96	69	2.7	89665	2	AC022403	Homo sapi
97	69	2.7	90727	2	AC015779	Homo sapi
98	69	2.7	93238	2	AC022713	Homo sapi
99	69	2.7	98261	2	HS1054C24	Human DNA
100	69	2.7	98261	2	HS1054C24	Human DNA

ALIGNMENTS

RESULT 1
LOCUS AX452880 2567 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 1 from Patent WO0242457.
ACCESSION AX452880
VERSION AX452880.1 GI:21712520
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Chang, H., Yang, W.P., Wu, Y., Whitney, G.S., Perez-Villar, J.J. and
Kanner, S.B.
TITLE Cloning and expression of human slap-2: a novel sn2/sh3
domain-containing human slap homologue having immune cell-specific
expression
JOURNAL Patent: WO 0242457-A 1 30-MAY-2002;
Bristol-Myers Squibb Co. (US)
FEATURES
source location/Qualifiers
1. 2567
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 611 a 741 c 666 g 549 t
ORIGIN

Query Match 100.0%; Score 2567; DB 6; Length 2567;
Best Local Similarity 100.0%; Pred. No 0;
Matches 2567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCACGGCTCCGGTGGAGCTGAGAGCTCCAGAGACCCACGCTGTGTCTCTGTGACAGA 60
Db 1 CCCACGGCTCCGGTGGAGCTGAGAGCTCCAGAGACCCACGCTGTGTCTCTGTGACAGA 60

Qy 61 GCTCAAGAGGCCCTGGGCTTCCCTCCGTGGCTGGCTGTGGAGGTTCCCAAGT 120
Db 61 GCTCAAGAGGCCCTGGGCTTCCCTCCGTGGCTGGCTGTGGAGGTTCCCAAGT 120
Qy 121 CCAGAAATCCCTAAGAGAGCTGGGAGCTATTCATCCCTGGTGTCAAACTGTGACTG 180
Db 121 CCAGAAATCCCTAAGAGAGCTGGGAGCTATTCATCCCTGGTGTCAAACTGTGACTG 180
Qy 181 CAGACAGATGCTAGCTACCCAAACCAACCTAGCTCTCCGTAAGATCTCCAGAGC 240
Db 181 CAGACAGATGCTAGCTACCCAAACCAACCTAGCTCTCCGTAAGATCTCCAGAGC 240
Qy 241 TGAGAGAGTCTGGGTCTTGGAGCCAAAGAGACATGGCAGACCTTCCAGAGGCCCA 300
Db 241 TGAGAGAGTCTGGGTCTTGGAGCCAAAGAGACATGGCAGACCTTCCAGAGGCCCA 300
Qy 301 AAGCCCTAAGCTGTCCAGCAGAGCATGCTCTCAGCAGAGCTGTCTTCCAGGCTTTG 360
Db 301 AAGCCCTAAGCTGTCCAGCAGAGCATGCTCTCAGCAGAGCTGTCTTCCAGGCTTTG 360
Qy 361 ATGACAAACCAATTTCCCTCATATGTCTGTAGTCTCTGAGGATCAATGGA 420
Db 361 ATGACAAACCAATTTCCCTCATATGTCTGTAGTCTCTGAGGATCAATGGA 420
Qy 421 AGCTGCTCAGCAGAAATCTCTGCTCAAGCCCAAGCTTGAATTCCTCTGCTCAAGGC 480
Db 421 AGCTGCTCAGCAGAAATCTCTGCTCAAGCCCAAGCTTGAATTCCTCTGCTCAAGGC 480
Qy 481 CAGGAGCTGTGACCAATGGAAGCAGAGAGCAGAGCAGAGCAGAGCAGAGCAGAGC 540
Db 481 CAGGAGCTGTGACCAATGGAAGCAGAGAGCAGAGCAGAGCAGAGCAGAGCAGAGC 540
Qy 541 TTCCGCGCAGTGGCCCGCGCAGAGCTGTGCTGAGACTCGGAGGACATTTGACATGCTC 600
Db 541 TTCCGCGCAGTGGCCCGCGCAGAGCTGTGCTGAGACTCGGAGGACATTTGACATGCTC 600
Qy 601 TGTGAGATGAGAGCTGGTGGAGCTGTGCTGAGAGCTTCCAGGAGATTAATC 660
Db 601 TGTGAGATGAGAGCTGGTGGAGCTGTGCTGAGAGCTTCCAGGAGATTAATC 660
Qy 661 CCCAGCTCCAGTGGTCCAAATGCTCCATGGTGGCTGTATGAGGCTTGAAGAGGAG 720
Db 661 CCCAGCTCCAGTGGTCCAAATGCTCCATGGTGGCTGTATGAGGCTTGAAGAGGAG 720
Qy 721 AAACAGAGGAACTGCTGTGTTACCTGGAGACCTGGAGGGCTTCTCTCATCCGGAG 780
Db 721 AAACAGAGGAACTGCTGTGTTACCTGGAGACCTGGAGGGCTTCTCTCATCCGGAG 780
Qy 781 AGCCAGACAGAGAGGCTTTACTCTGTGACGTCCGCTCAGCGCTGCAATCCCTGG 840
Db 781 AGCCAGACAGAGAGGCTTTACTCTGTGACGTCCGCTCAGCGCTGCAATCCCTGG 840
Qy 841 GACCGATCAGACATACAGATCCATGCTGCTGAGAGGCTGCTGATCCATCCACCG 900
Db 841 GACCGATCAGACATACAGATCCATGCTGCTGAGAGGCTGCTGATCCATCCACCG 900
Qy 901 CGCTCACTTCCCTCACTCAGGCTGCTGAGCACTTACTTCTGAGCTGCGGATGAC 960
Db 901 CGCTCACTTCCCTCACTCAGGCTGCTGAGCACTTACTTCTGAGCTGCGGATGAC 960
Qy 961 ATGCTGCTTACTAAGAGAGCTGTGCTGCTGAGAGGCTGGGCTTCCCTGGAGAG 1020
Db 961 ATGCTGCTTACTAAGAGAGCTGTGCTGCTGAGAGGCTGGGCTTCCCTGGAGAG 1020
Qy 1021 GATATACCTTACTGCTGAGCTGTGAGAGACACCACTCACTGAGAAAGCTGAGCAGC 1080
Db 1021 GATATACCTTACTGCTGAGCTGTGAGAGACACCACTCACTGAGAAAGCTGAGCAGC 1080
Qy 1081 TTCCCTCTTCTTCTGAGAGCTGACAGAGGAGAGCTTCTCTGAGAGAGGCTTCCG 1140
Db 1081 TTCCCTCTTCTTCTGAGAGCTGACAGAGGAGAGCTTCTCTGAGAGAGGCTTCCG 1140

Qy 1141 GAGTCCCTCAGCTTCTAATCAGCTGAAATGACGAGCTGTCTCTTTGGATGATGCTAG 1200
 Db 1141 GAGTCCCTCAGCTTCTAATCAGCTGAAATGACGAGCTGTCTCTTTGGATGATGCTAG 1200
 Qy 1201 GCCCAAGAGAGGCCCCAAAAGGAAAACAAAGCTGCACTAGAACCCCAATTCAGCCT 1260
 Db 1201 GCCCAAGAGAGGCCCCAAAAGGAAAACAAAGCTGCACTAGAACCCCAATTCAGCCT 1260
 Qy 1261 CCGGGGCAACCCAGAGAGGCTGTGCACTCAGGGAGGAGGCTGGGACACAGAGTGC 1320
 Db 1261 CCGGGGCAACCCAGAGAGGCTGTGCACTCAGGGAGGAGGCTGGGACACAGAGTGC 1320
 Qy 1321 ATCTAGAGGTCCCACTGTATCCCTGCTCTTTCTCTCTTACGCTTAGAGTACCTACT 1380
 Db 1321 ATCTAGAGGTCCCACTGTATCCCTGCTCTTTCTCTCTTACGCTTAGAGTACCTACT 1380
 Qy 1381 TCGCTTCAGTGGCAATGATCCCACTGCACTGAGTGGAGTGCAGAGAAAGTGGGACC 1440
 Db 1381 TCGCTTCAGTGGCAATGATCCCACTGCACTGAGTGGAGTGCAGAGAAAGTGGGACC 1440
 Qy 1441 AGGGCCAGAGGTTCCAAAAGAGAAATAGCCTCTGGGGGGTCTGACCTAGTTAGTTCTTG 1500
 Db 1441 AGGGCCAGAGGTTCCAAAAGAGAAATAGCCTCTGGGGGGTCTGACCTAGTTAGTTCTTG 1500
 Qy 1501 AGTTTGGGGTTTCCAGTACCATCTGATGCGCTGCGCTGTGAGCCCACTTCTACATCCC 1560
 Db 1501 AGTTTGGGGTTTCCAGTACCATCTGATGCGCTGCGCTGTGAGCCCACTTCTACATCCC 1560
 Qy 1561 ACCATTAAACAGAGCCCAACCAAGAGTAAACAAACCCCTAGTCAACAGAGAAAGTCA 1620
 Db 1561 ACCATTAAACAGAGCCCAACCAAGAGTAAACAAACCCCTAGTCAACAGAGAAAGTCA 1620
 Qy 1621 TTTTCAGAAATCTCAAGTCTGTGAGACCAACCACTTACCTAGAGAGTGAAGCTCT 1680
 Db 1621 TTTTCAGAAATCTCAAGTCTGTGAGACCAACCACTTACCTAGAGAGTGAAGCTCT 1680
 Qy 1681 GGCCTAAGAGGAAAGGAAAGAGTGAATGATGTCTTACCTAGAGAGTCTTGATG 1740
 Db 1681 GGCCTAAGAGGAAAGGAAAGAGTGAATGATGTCTTACCTAGAGAGTCTTGATG 1740
 Qy 1741 TCCAGGCTTATGACCTCCAGACCAAGAGAAAGCTTCCGAGAGTCAAGCTGCTCA 1800
 Db 1741 TCCAGGCTTATGACCTCCAGACCAAGAGAAAGCTTCCGAGAGTCAAGCTGCTCA 1800
 Qy 1801 ATGTCCTCCATGAGCAACAGCCCAAGCTCTTTTCTTTTGTGAGACGAGTCTT 1860
 Db 1801 ATGTCCTCCATGAGCAACAGCCCAAGCTCTTTTCTTTTGTGAGACGAGTCTT 1860
 Qy 1861 GGCCTGTTGACCATGAGTGAATGAGCAATGAGTCACTGCAACCTCCATCTC 1920
 Db 1861 GGCCTGTTGACCATGAGTGAATGAGTGAATGAGTCACTGCAACCTCCATCTC 1920
 Qy 1921 CTGATTCAAACAAATTCCTCTGCTCCTCCTCCAGATAGCTGGATTAACGAGTAC 1980
 Db 1921 CTGATTCAAACAAATTCCTCTGCTCCTCCTCCAGATAGCTGGATTAACGAGTAC 1980
 Qy 1981 CACCATGCTGCTAATTTTGTATTTTGTATGAGCAATGAGGTTTACCAATTTGGCC 2040
 Db 1981 CACCATGCTGCTAATTTTGTATTTTGTATGAGCAATGAGGTTTACCAATTTGGCC 2040
 Qy 2041 AGGCTGTTGAGAACTCTGACCTCAGGTATCCCACTTGGCTCCCAAGTCTG 2100
 Db 2041 AGGCTGTTGAGAACTCTGACCTCAGGTATCCCACTTGGCTCCCAAGTCTG 2100
 Qy 2101 GGATTCAGAGTGTAGAGCAACGCAACCAAGCTGCTCAGATCTCTATTTTCTTGG 2160
 Db 2101 GGATTCAGAGTGTAGAGCAACGCAACCAAGCTGCTCAGATCTCTATTTTCTTGG 2160
 Qy 2161 GCTTACCAATTCCTAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 Db 2161 GCTTACCAATTCCTAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 Qy 2221 CTTAAGTCAAGCACTGAGTGAAGGCAAGGCACTTCAAGTCTGCTGCTGCTGCTGCTG 2280

Db 2221 CTTAAGTCAAGCACTGAGTGAAGGCAAGGCACTTCAAGTCTGCTGCTGCTGCTG 2280
 Qy 2281 GGTGTAAAGCCCTCTCTCCAAATGCAAGAGCAACAGGCTTACCAATTCAGC 2340
 Db 2281 GGTGTAAAGCCCTCTCTCCAAATGCAAGAGCAACAGGCTTACCAATTCAGC 2340
 Qy 2341 CTTGATTTTCTGCTGCT 2400
 Db 2341 CTTGATTTTCTGCTGCT 2400
 Qy 2401 AGGAG 2460
 Db 2401 AGGAG 2460
 Qy 2461 CCAAGCTTCAAGTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
 Db 2461 CCAAGCTTCAAGTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
 Qy 2521 AGATTGATTTTCAAGTGTGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2567
 Db 2521 AGATTGATTTTCAAGTGTGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2567

RESULT 2
 AK025645
 LOCUS
 DEFINITION
 Homo sapiens cDNA: FLJ21992 fib, clone HEP06554.
 AK025645
 ACCESSION
 AK025645.1 GI:10438227
 VERSION
 cligo capring: fib (full insert sequence)
 KEYWORDS
 Homo sapiens
 Hepatoma cell_line:hep02 cDNA to mRNA, clone_lib:HEP
 clone:HEP06554.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS
 Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
 Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
 Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
 NEBO human cDNA sequencing project

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure Analysis, Human
 Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
 Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)
 NEBO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction, 5'- 6.3'-end one pass sequencing; Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

COMMENT
 FEATURES
 source
 Location/Qualifiers
 1..2415
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEP06554"
 /cell_line="HepG2"
 /cell_type="hepatoma"
 /clone_lib="HEP"
 /note="Cloning vector pME18SF13"
 61..846
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB15201.1"
 /db_xref="GI:10438228"
 /translation="MGSLSRKRKSLPSPSSSSVQGGPVTMEARSKATPAVALGSPF

CDS

/note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB15201.1"
 /db_xref="GI:10438228"
 /translation="MGSLSRKRKSLPSPSSSSVQGGPVTMEARSKATPAVALGSPF

BASE COUNT	ORIGIN
599 a	662 c 618 g 536 t

Query Match	72.5%	Score 1860	DB 9	Length 2415
Best Local Similarity	99.8%	Pred. No. 0		
Matches 2130; Conservative	0	Mismatches 3	Indels 1	Gaps 1

415 ATGGAAAGCTCCCGACAGAAAGAAATCTGCGCAGAGCCAGCTTGAGTTCCTGTC 474
 Db 61 ATGGAAAGCTCCCGACAGAAAGAAATCTGCGCAGAGCCAGCTTGAGTTCCTGTC 120
 Oy 475 CAAAGCCAGAGACCTGTGACCATGAAACAAGAGAAAGCAAGGCCAACGCTGGCCCTG 534
 Db 121 CAAAGCCAGAGACCTGTGACCATGAAACAAGAGAAAGCAAGGCCAACGCTGGCCCTG 180
 Oy 535 GGCAGTTTCCCGACAGTGGCCCGACGACCTGTGCTGAAGCTCGGGAGGCAATTAGC 594
 Db 181 GGCAGTTTCCCGACAGTGGCCCGACGACCTGTGCTGAAGCTCGGGAGGCAATTAGC 240
 Oy 595 ATGCTCTTGAAGGATGAGACTGTGTGACGGTGTCTGTAGTCTCAGGCAAGAGAT 654
 Db 241 ATGCTCTTGAAGGATGAGACTGTGTGACGGTGTCTGTAGTCTCAGGCAAGAGAT 300
 Oy 655 AACATCCCGACGGTCACTGAGCCAAAGTCTCCATGGGTGGCTGATAGAGGACTAGC 714
 Db 301 AACATCCCGACGGTCACTGAGCCAAAGTCTCCATGGGTGGCTGATAGAGGACTAGC 360
 Oy 715 AGGAGAAAGCAGAGAGAACTGTGTTTAACTGGGAAACCTGAGAGGGCTTCTCATC 774
 Db 361 AGGAGAAAGCAGAGAGAACTGTGTTTAACTGGGAAACCTGAGAGGGCTTCTCATC 420
 Oy 775 CGGAGAGCCAAACAGAGAGAGGCTCTTACTCTGTGATGCGCCTCAAGCCGCTGCA 834
 Db 421 CGGAGAGCCAAACAGAGAGAGGCTCTTACTCTGTGATGCGCCTCAAGCCGCTGCA 480
 Oy 835 TCTGTGGACCGGATCAAGCATACAGATTCATCTGCTTGAACAAATGGCTGTATCATC 894
 Db 481 TCTGTGGACCGGATCAAGCATACAGATTCATCTGCTTGAACAAATGGCTGTATCATC 540
 Oy 895 TCAACCGGCTCACTTCCTCTCACTCAGAGCCCTGTGTGACATTAATCTGAGCTGAGC 954
 Db 541 TCAACCGGCTCACTTCCTCTCACTCAGAGCCCTGTGTGACATTAATCTGAGCTGAGC 600
 Oy 955 GATGACATCTGCTGCTACTAGAGAGCCCTGTGTCTGCAAGAGGCTGGCCGCTCCCT 1014
 Db 601 GATGACATCTGCTGCTACTAGAGAGCCCTGTGTCTGCAAGAGGCTGGCCGCTCCCT 660
 Oy 1015 GGCAGAGATATACCCCTACTGCTGTGACTGTGCAGAGAGACCAACACTCAACTGGAAAGAGCTG 1074
 Db 661 GGCAGAGATATACCCCTACTGCTGTGACTGTGTGAGAGAGACCAACACTCAACTGGAAAGAGCTG 720
 Oy 1075 GACAGCTCCCTCTGTTTCTGAAAGCTGCAAGGGAGAGATCTCTTCACTGAGAGGT 1134
 Db 721 GACAGCTCCCTCTGTTTCTGAAAGCTGCAAGGGAGAGATCTCTTCACTGAGAGGT 780
 Oy 1135 CTCGCGGAGTCCCTCAGCTTCTATCATCAGCTGAATGACGAGGCTGCTCTTTGGATGAT 1194
 Db 781 CTCGCGGAGTCCCTCAGCTTCTATCATCAGCTGAATGACGAGGCTGCTCTTTGGATGAT 840
 Oy 1195 GCTTAGGCCCAAAAGAGAGGCAAAAGGAAACAAAGGCTGCACACTGAAACCCCAATT 1254
 Db 841 GCTTAGGCCCAAAAGAGAGGCAAAAGGAAACAAAGGCTGCACACTGAAACCCCAATT 900
 Oy 1255 CAGGCTCTTGGGACCCCAAGAGGCAAGGCTGTGCACTCAGGAGAGGAGGGTGGGACAG 1314
 Db 901 CAGGCTCTTGGGACCCCAAGAGGCAAGGCTGTGCACTCAGGAGAGGAGGGTGGGACAG 960
 Oy 1315 AGGTGCACTTAGGGTCCCACTGTACCTTGTCTTCTCTTGAACCTTAGAAGTCA 1374

Db	961	AGGTGCAATCAGAGGCTCCACCTGACCTTGTCTTTTCCTCTTACGCCCTTGAAGTCA	1070
Oy	1375	CCGACTTCTTCCAGTGGCAATGATCCCACTTGCAGACTCTAGTGGATGACGAAAGT	1434
Db	1021	CCATCTTCTTCCAGTGGCAATGATCCCACTTGCAGACTCTAGTGGATGACGAAAGT	1080
Oy	1435	GGGACAGAGGCGAGGGTTCAAAAAGAAATAGCTCTCTGGGGGCTTGACTTAATTG	1494
Db	1081	GGGACAGAGGCGAGGGTTCAAAAAGAAATAGCTCTCTGGGGGCTTGACTTAATTG	1140
Oy	1495	TTCTTGATTTGGGGTTTCCAGTACATGTGGATGCCCTGCTGTGAGCCCATTTAC	1554
Db	1141	TTCTTGATTTGGGGTTTCCAGTACATGTGGATGCCCTGCTGTGAGCCCATTTAC	1200
Oy	1555	ATCCCAACATTACAGAGGCCCAACCAAGATGAAATACCCCTAGAGTCAAGACA	1614
Db	1201	ATCCCAACATTACAGAGGCCCAACCAAGATGAAATACCCCTAGAGTCAAGACA	1260
Oy	1615	AAGTCATTTTCAGAAATCTCAAGTCTGTTGAGACCAACACATACCTCAGAGTAG	1674
Db	1261	AAGTCATTTTCAGAAATCTCAAGTCTGTTGAGACCAACACATACCTCAGAGTAG	1320
Oy	1675	GACTGTGCTTAAGAGGAAAGAAAGCTGAAATGATCTTACCGTGAAGCAAGATCT	1734
Db	1321	GACTGTGCTTAAGAGGAAAGAAAGCTGAAATGATCTTACCGTGAAGCAAGATCT	1380
Oy	1735	GGATGTCCAGGCTCTAATGTACCTCCAGAGCAAGAAAGACTTCCGACATGTAGT	1794
Db	1381	GGATGTCCAGGCTCTAATGTACCTCCAGAGCAAGAAAGACTTCCGACATGTAGT	1440
Oy	1795	CCTCAAATGCCCCCATTTGAGAAACAAGCCCAAGCTCTTTTCTTTTTTTGAAGCG	1854
Db	1441	CCTCAAATGCCCCCATTTGAGAAACAAGCCCAAGCTCTTTTCTTTTTTGAAGCG	1499
Oy	1855	AGCTTGGCCCTGTTTGGCTCCTAGTGGATGTGCAATGTGCAAGATTCAGCTCACTGCAACCTC	1914
Db	1500	AGCTTGGCCCTGTTTGGCTCCTAGTGGATGTGCAATGTGCAAGATTCAGCTCACTGCAACCTC	1559
Oy	1915	CATCTCCGGAATTCAAACAATTTCCCTGCTCAGCGCTCCAGATATGCTGGGATTAAGGCG	1974
Db	1560	CATCTCCGGAATTCAAACAATTTCCCTGCTCAGCGCTCCAGATATGCTGGGATTAAGGCG	1619
Oy	1975	GTACACACCAATCCTGCGCTAATTTTTTGTATTTTGTATAGTGAATGGGGTTTCAACCA	2034
Db	1620	GTACACACCAATCCTGCGCTAATTTTTTGTATTTTGTATAGTGAATGGGGTTTCAACCA	1679
Oy	2035	TTGGCAGAGCGGTGTCCAACTCTGACCTCAGGTGATTCACCAACCTTGGCTCCGCAAA	2094
Db	1680	TTGGCAGAGCGGTGTCCAACTCTGACCTCAGGTGATTCACCAACCTTGGCTCCGCAAA	1739
Oy	2095	GTGCTGGGATTAAGAGTGTGAGCCAGGCAACCAGCTAGCTCTCAGATCTTAATTTTCA	2154
Db	1740	GTGCTGGGATTAAGAGTGTGAGCCAGGCAACCAGCTAGCTCTCAGATCTTAATTTTCA	1799
Oy	2155	TTTTGTGCTTAACAATTCCTATGACAATGGCCCTTGCATCTCTGTGCGAATTAATAAAA	2214
Db	1800	TTTTGTGCTTAACAATTCCTATGACAATGGCCCTTGCATCTCTGTGCGAATTAATAAAA	1854
Oy	2215	ACACCTTTAATGCTTAGACAATCAGTAGAGCCAGGCACTCAGTGTGGGCAAGGCA	2274
Db	1860	ACACCTTTAATGCTTAGACAATCAGTAGAGCCAGGCACTCAGTGTGGGCAAGGCA	1919
Oy	2275	TCAAGAAAGTGTAAAGCCCTCTCTCAATGTCAAGACGAGACAACAAGCTTACACAAA	2333
Db	1920	TCAAGAAAGTGTAAAGCCCTCTCTCAATGTCAAGACGAGACAACAAGCTTACACAAA	1979
Oy	2335	TCCAGCCTTGAATTTCTCTGCTGCCTCAATAAAGAAAGATCTGCTGATTCGCTAA	2394
Db	1980	TCCAGCCTTGAATTTCTCTGCTGCCTCAATAAAGAAAGATCTGCTGATTCGCTAA	2038
Oy	2395	GGGATCAAGAGAGAAAGAAAGAGATGGGGTGGAGAGCAACCCCTCAGATCTCTTAC	2454
Db	2040	GGGATCAAGAGAGAAAGAAAGAGATGGGGTGGAGAGCAACCCCTCAGATCTCTTAC	2099


```

/codon_start=3
/evidence=not_experimental
/product="dj977b1.6.2 (KIA0964 (ortholog of rat
PSD-95/SAP90-associated protein 4))"
/protein_id="CAC16960.1"
/db_xref="GI:11323333"
/db_xref="SPRMBL:09H138"
/translation="SSCLVAYKATPPVPVPRRTTSKPFISVTVOGSTESADPTLDSOD
HKEVTSQSLNSDLSSTRPPTVTRGPAPEKPEPPPHALKEGGLTTS
ESHPEAPKRLKSSIGIERTRRNGSHLSEDPGPAIDVMAPESSSVSHMSRRD
LSDTQDAPNSCKSBSRLPCTPHNLSISIDAGRPAPKIAQIKRLNSYGDNDPA
LEASLPDPVPLETSSSPAPPAQPCACRQGVFLKLAETRLKELCCOMKET
KENNISEVLGKVLASVGAOLMSQFOFRLCEONLNDAPRLEODLAGWDL
LQSLIEDISMKPDELHYLKANSWQLVETPEKKEKPPVPEKPKAKSPVSDKA
SDADKORARAKRLAAKRAASVROKATESADSIETVPEAQTR"
/contig="7423..7784.10306..10394.10917..11329.34968..35059,
35569..36724.37531..37749"
/gene="KIA0964"
/note="isoform 1
continues in dj996C2 (AL445705) and bx2182L9 (AL390374)
match: proteins: Tr:014488 Tr:014664 Tr:014489 Tr:P97839
Tr:014490 Tr:P97836 Tr:P97837 Tr:054773 Tr:P97841
Tr:P78335"
/codon_start=3
/evidence=not_experimental
/product="dj977b1.6.1 (KIA0964 (ortholog of rat
PSD-95/SAP90-associated protein 4))"
/protein_id="CAB75170.1"
/db_xref="GI:6983730"
/db_xref="SPRMBL:09H137"
/translation="SSCLVAYKATPPVPVPRRTTSKPFISVTVOGSTESADPTLDSOD
HKEVTSQSLNSDLSSTRPPTVTRGPAPEKPEPPPHALKEGGLTTS
ESHPEAPKRLKSSIGIERTRRNGSHLSEDPGPAIDVMAPESSSVSHMSRRD
LSDTQDAPNSCKSBSRLPCTPHNLSISIDAGRPAPKIAQIKRLNSYGDNDPA
LEASLPDPVPLETSSSPAPPAQPCACRQGVFLKLAETRLKELCCOMKET
KENNISEVLGKVLASVGAOLMSQFOFRLCEONLNDAPRLEODLAGWDL
LQSLIEDISMKPDELHYLKANSWQLVETPEKKEKPPVPEKPKAKSPVSDKA
SDADKORARAKRLAAKRAASVROKATESADSIETVPEAQTR"
10065..10112
/note="6 copies 8 mer cgcctgc 79% conserved"
11537..11753
/note="WIR repeat: matches 11..236 of consensus"
11767..11782
/note="L2 repeat: matches 2692..2708 of consensus"
11783..12083
/note="AluX repeat: matches 1..300 of consensus"
12084..12121
/note="L2 repeat: matches 2661..2692 of consensus"
12631..12714
/note="HAU1 repeat: matches 1589..1665 of consensus"
13077..13367
/note="AluDo repeat: matches 3..290 of consensus"
13440..13742
/note="AluX repeat: matches 1..304 of consensus"
13932..14197
/note="L1MB7 repeat: matches 5915..6173 of consensus"
14198..14512
/note="AluY repeat: matches 1..307 of consensus"
14513..14756
/note="L1MB7 repeat: matches 5650..5915 of consensus"
14789..15705
/note="L1MB4 repeat: matches 5162..6185 of consensus"
15706..16004
/note="AluSg repeat: matches 1..298 of consensus"
16005..16060
/note="L1MB4 repeat: matches 5111..5162 of consensus"

Query Match 52.7%; Score 1352; DB 9; Length 145068;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1472; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1076 ACAGCTCCCTGTTTCTGAGCTCCAGGAGAGATCTTCTCAGTAGAGGCT 1135
|||||

```

```

DB 124708 ACAGCTCCCTGTTTCTGAGCTCCAGGAGAGATCTTCTCAGTAGAGGCT 124649
QY 1136 TCCGGAGATCCCTCAGCTTCAATCAGCTGAATGACAGAGCTGCTCTTGTGATGATG 1195
DB 124648 TCCGGAGATCCCTCAGCTTCAATCAGCTGAATGACAGAGCTGCTCTTGTGATGATG 124589
QY 1196 CTAAGGCCCAAGAGAGAGGCAAGAGGAAACCAAGGCTGCAACCTGAACCCCAATTC 1255
DB 124588 CTAAGGCCCAAGAGAGAGGCAAGAGGAAACCAAGGCTGCAACCTGAACCCCAATTC 124529
QY 1256 AGCCTCTGGGCAACCCAGAGCAAGGCTGTGCACTCAGGAGGAGGCTGGGACACAGA 1315
DB 124528 AGCCTCTGGGCAACCCAGAGCAAGGCTGTGCACTCAGGAGGAGGCTGGGACACAGA 124469
QY 1316 GGTGCACTCAGGAGCCCACTGACCTGCTCTTCTCTCTAGAGCTAGAGTCA 1375
DB 124468 GGTGCACTCAGGAGCCCACTGACCTGCTCTTCTCTCTAGAGCTAGAGTCA 124409
QY 1376 CTACTTCTTCCAGTGCATGATCCCACTGCGACCTTAGTGCGAGTGACAGAGAGTGTG 1435
DB 124408 CTACTTCTTCCAGTGCATGATCCCACTGCGACCTTAGTGCGAGTGAGAGAGTGTG 124349
QY 1436 GAGCAGGCGCAGGCTTCCAAAAGAGAAATAGCCTCTGGGGGGTGTGACCTAGTAGT 1495
DB 124348 GAGCAGGCGCAGGCTTCCAAAAGAGAAATAGCCTCTGGGGGGTGTGACCTAGTAGT 124289
QY 1496 TCTTGAGTTGGGGTTTCCAGTACATCTGAGTCCCTGCTGTTGAGCCCATTTCA 1555
DB 124288 TCTTGAGTTGGGGTTTCCAGTACATCTGAGTCCCTGCTGTTGAGCCCATTTCA 124229
QY 1556 TCCCAACATTAACAGAGGCCCAACCAAGAGTGAAGAACCCCTAGAGTCAAGAGAA 1615
DB 124228 TCCCAACATTAACAGAGGCCCAACCAAGAGTGAAGAACCCCTAGAGTCAAGAGAA 124169
QY 1616 AGTCAATTTCAAGAAATCTCAAGTCTGTTGAGACCAACACATACCTCAGAGGTAG 1675
DB 124168 AGTCAATTTCAAGAAATCTCAAGTCTGTTGAGACCAACACATACCTCAGAGGTAG 124109
QY 1676 ACTGTGCGCTGAAGAGAGAAAGAGAGTGAATGATCTTACCTGACAGAGATCTTG 1735
DB 124108 ACTGTGCGCTGAAGAGAGAAAGAGAGTGAATGATCTTACCTGACAGAGATCTTG 124049
QY 1736 GATGTGCAAGCTCTATGTGACCTCCAGAGAAAGAAAGCTTGGACAGTGTAGCTC 1795
DB 124048 GATGTGCAAGCTCTATGTGACCTCCAGAGAAAGAAAGCTTGGACAGTGTAGCTC 123989
QY 1796 CTCAATGTCCTCCATTTGAGAGCAACAGAGCCAGCTTTTCTTTTGTGAGACGGA 1855
DB 123988 CTCAATGTCCTCCATTTGAGAGCAACAGAGCCAGCTTTTCTTTTGTGAGACGGA 123930
QY 1856 GTCTTGCTGCTGTCCTATGCTGAGAGTGCATGAGCATCTGACCTAGTCAACTCC 1915
DB 123930 GTCTTGCTGCTGTCCTATGCTGAGAGTGCATGAGCATCTGACCTAGTCAACTCC 123870
QY 1916 ATCTCTGATTTCAAAATTTCTCTGCTCAGGCTCAGAAATAGCTGAGTTACAGGCG 1975
DB 123870 ATCTCTGATTTCAAAATTTCTCTGCTCAGGCTCAGAAATAGCTGAGTTACAGGCG 123810
QY 1976 TACACCAACAGCTGCTGCTAATTTTGTATTTTGTATTTTGTATTTGATGACATGGGTTTCA 2035
DB 123810 TACACCAACAGCTGCTGCTAATTTTGTATTTTGTATTTTGTATTTGATGACATGGGTTTCA 123750
QY 2036 TGGCAGAGTGTGTCGACATCTGACCTCAGAGTGCACCACTTGGCTGCCAAG 2095
DB 123750 TGGCAGAGTGTGTCGACATCTGACCTCAGAGTGCACCACTTGGCTGCCAAG 123690
QY 2096 TGTGAGGATTAAGGTGAGCGACAGCGACCAAGCTGATCTGAGATCTTATTTATT 2155
DB 123690 TGTGAGGATTAAGGTGAGCGACAGCGACCAAGCTGATCTGAGATCTTATTTATT 123630
QY 2156 TTGTGCTTACATTTCTTACACTGAGCACTGAGCTTGCATCTTGTGCGCAATTAATAATA 2215
DB 123630 TTGTGCTTACATTTCTTACACTGAGCACTGAGCTTGCATCTTGTGCGCAATTAATAATA 123570

```


QY 2216 CACCTTTAAGTCTAGACACCTGACGTAGAGCCAGGACCTGACGTGCGG CAGGGGCT 2275
 Db 123569 CACCTTTAAGTCTAGACACCTGACGTAGAGCCAGGACCTGACGTGCGG CAGGGGCT 123510
 QY 2276 CAGAAAGTCTAGAGCCCTCTCTCCAAATGCGAAGAGGACCAAGCTTCAACCAAT 2335
 Db 123509 CAGAAAGTCTAGAGCCCTCTCTCCAAATGCGAAGAGGACCAAGCTTCAACCAAT 123450
 QY 2336 CAGAGCCCTTGAATTTCCCTGCTGCTCCATTAACGAAAGAGGTCTGGAATCCGCTAAG 2395
 Db 123449 CAGAGCCCTTGAATTTCCCTGCTGCTCCATTAACGAAAGAGGTCTGGAATCCGCTAAG 123390
 QY 2396 GATCAGGAGAGGAGAAAGAGGATGGGGTGGAGAGCAACCCCTCCAGTCTCTACT 2455
 Db 123389 GATCAGGAGAGGAGAAAGAGGATGGGGTGGAGAGCAACCCCTCCAGTCTCTACT 123330
 QY 2456 GGTTCCTCAAGCTACAGGTGGGGTGGAAAGCTTATCAGATATCATCAAGGTTCTCA 2515
 Db 123329 GGTTCCTCAAGCTACAGGTGGGGTGGAAAGCTTATCAGATATCATCAAGGTTCTCA 123270
 QY 2516 ATTAAAGATTGATTATTCAGATATGTGAAA 2548
 Db 123269 ATTAAAGATTGATTATTCAGATATGTGAAA 123237

 RESULT 4
 AX443133 1183 bp DNA linear PAT 02-JUL-2002
 LOCUS AX443133
 DEFINITION Sequence 74 from Patent WO216599.
 ACCESSION AX443133
 VERSION AX443133.1 GI:21690555
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 AUTHORS
 REFERENCES
 1
 TITLE
 JOURNAL
 Curagen Corporation (US); COR THERAPEUTICS, INC. (US)
 Location/Qualifiers
 source 1. 1183
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 251 a 359 c 333 g 240 t
 ORIGIN
 Query Match 44.1%; Score 1132; DB 6; Length 1183;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 QY 18 AGCTAGAGCTCAAGAGCCACGACCTGCTCTGTCAGACAGAGCTCAAGAGGCGCTGAG 77
 Db 1 AGCTAGAGCTCAAGAGCCACGACCTGCTCTGTCAGACAGAGCTCAAGAGGCGCTGAG 60
 QY 78 CTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 137
 Db 61 CTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
 QY 138 CATGGGAGCTGATCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 197
 Db 121 CATGGGAGCTGATCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 QY 198 ACCCAACCAACCTGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 257
 Db 181 ACCCAACCAACCTGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 QY 258 TCCTAGAGCAAGAGCACTGCGAGACTTCCAGAAAGGCGCCCAAGCCCTTAACCTGTCCA 317

Db 241 TCCTAGAGCAAGAGCACTGCGAGACTTCCAGAAAGGCGCCCAAGCCCTTAACCTGTCCA 300
 QY 318 GCGAGAGATGCTCTCAGCAGAGCTGTCTTCCCAAGCCTTTGATGCAACCAATTTCC 377
 Db 301 GCGAGAGATGCTCTCAGCAGAGCTGTCTTCCCAAGCCTTTGATGCAACCAATTTCC 360
 QY 378 CTCGATGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 437
 Db 361 CTCGATGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 QY 438 AAAATCTCTGCAAGCCCAAGCTTGAATTTCTTGTCCAGAGGACCTGTGACAT 497
 Db 421 AAAATCTCTGCAAGCCCAAGCTTGAATTTCTTGTCCAGAGGACCTGTGACAT 480
 QY 498 GGAACAGAGAGAGCAAGCCCAAGCTGAGCGGCGCTGAGAGTTTCCGAGAGTGGCCC 557
 Db 481 GGAACAGAGAGAGCAAGCCCAAGCTGAGCGGCGCTGAGAGTTTCCGAGAGTGGCCC 540
 QY 558 GGCAGAGCTGCTGCTGAGACTGAGGAGCATTGACATCGTCTGAGAGATGAGACTG 617
 Db 541 GGCAGAGCTGCTGCTGAGACTGAGGAGCATTGACATCGTCTGAGAGATGAGACTG 600
 QY 618 GTGACGCTGCTGCTGAGACTGAGAGATTAACATCCCGAGCTGCCAGCTGAC 677
 Db 601 GTGACGCTGCTGCTGAGACTGAGAGATTAACATCCCGAGCTGCCAGCTGAC 660
 QY 678 CAAAGTCTCCAGAGGCTGCTGATGAGGCTGAGCAGAGGAGAAACAGAGAACTGCT 737
 Db 661 CAAAGTCTCCAGAGGCTGCTGATGAGGCTGAGCAGAGGAGAAACAGAGAACTGCT 720
 QY 738 GTTGTACCTGAGAACCTGAGAGGCGCTTCTCATCCGAGAGAGCCAGAGAGAG 797
 Db 721 GTTGTACCTGAGAACCTGAGAGGCGCTTCTCATCCGAGAGAGCCAGAGAGAG 780
 QY 798 CTCTTACTCTCTGCTCAGTCCGCTTACGCCCTGCTGATCTGAGAACCGGATCAGAACTA 857
 Db 781 CTCTTACTCTCTGCTCAGTCCGCTTACGCCCTGCTGATCTGAGAACCGGATCAGAACTA 840
 QY 858 CAGATTCACCTGCTTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 917
 Db 841 CAGATTCACCTGCTTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 QY 918 ACTTCAGAGCCCTGCTGAGACATTAATCTGAGCTGGGAGATGACATCTGCTGCTGCTGCT 977
 Db 901 ACTTCAGAGCCCTGCTGAGACATTAATCTGAGCTGGGAGATGACATCTGCTGCTGCTGCT 960
 QY 978 GAGAGCCCTGCTGCTGAGAGGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1037
 Db 961 GAGAGCCCTGCTGCTGAGAGGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 QY 1038 GACTGCTGAGAGCACTCACTTACTGGAAGAGCTGAGCAGCTTCTCTGCTGCTGCTGCTGCT 1097
 Db 1081 GACTGCTGAGAGCACTCACTTACTGGAAGAGCTGAGCAGCTTCTCTGCTGCTGCTGCTGCT 1140
 QY 1158 CATCAGCTGATGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 Db 1141 CATCAGCTGATGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1183

 RESULT 5
 AX443135/c 1183 bp DNA linear PAT 02-JUL-2002
 LOCUS AX443135
 DEFINITION Sequence 76 from Patent WO216599.
 ACCESSION AX443135
 VERSION AX443135.1 GI:21690556
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L7115
Center clone name: 712.N.14
----- Summary Statistics -----
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phred; version 0.960731
Consensus quality: 125577 bases at least Q40
Consensus quality: 135703 bases at least Q30
Consensus quality: 139593 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 142233; sum-of-contigs
Quality coverage: 2.6 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1214: contig of 1214 bp in length
* 1215 1314: gap of 100 bp
* 1315 2673: contig of 1359 bp in length
* 2674 2773: gap of 100 bp
* 2774 4520: contig of 1747 bp in length
* 4521 4620: gap of 100 bp
* 4621 5961: contig of 1341 bp in length
* 5962 6061: gap of 100 bp
* 6062 7719: contig of 1658 bp in length
* 7720 7819: gap of 100 bp
* 7820 9799: contig of 1980 bp in length
* 9800 9899: gap of 100 bp
* 9900 11434: contig of 1535 bp in length
* 11435 11534: gap of 100 bp
* 11535 14382: contig of 2848 bp in length
* 14383 14482: gap of 100 bp
* 14483 17146: contig of 2666 bp in length
* 17149 17248: gap of 100 bp
* 17249 19239: contig of 1991 bp in length
* 19240 19339: gap of 100 bp
* 19340 21102: contig of 1763 bp in length
* 21103 21202: gap of 100 bp
* 21203 23371: contig of 2169 bp in length
* 23372 23471: gap of 100 bp
* 23472 25782: contig of 2311 bp in length
* 25783 25882: gap of 100 bp
* 25883 28824: contig of 2942 bp in length
* 28825 28924: gap of 100 bp
* 28925 31619: contig of 2695 bp in length
* 31620 31719: gap of 100 bp
* 31720 34679: contig of 2960 bp in length
* 34680 34779: gap of 100 bp
* 34780 37527: contig of 2748 bp in length
* 37528 37627: gap of 100 bp
* 37628 40146: contig of 2519 bp in length
* 40147 40246: gap of 100 bp
* 40247 43743: contig of 3497 bp in length
* 43744 43843: gap of 100 bp
* 43844 47133: contig of 3290 bp in length
* 47134 47233: gap of 100 bp
* 47234 51023: contig of 3790 bp in length
* 51024 51123: gap of 100 bp
* 51124 54735: contig of 3812 bp in length
* 54736 55035: gap of 100 bp
* 55036 59553: contig of 4518 bp in length
* 59554 59653: gap of 100 bp

```

```

FEATURES
source
* 59654 64428: contig of 4775 bp in length
* 64429 64528: gap of 100 bp
* 64529 69211: contig of 4683 bp in length
* 69212 69311: gap of 100 bp
* 69312 72901: contig of 3590 bp in length
* 72902 73001: gap of 100 bp
* 73002 76724: contig of 3723 bp in length
* 76725 76824: gap of 100 bp
* 76825 81179: contig of 4355 bp in length
* 81180 81279: gap of 100 bp
* 81280 87009: contig of 5730 bp in length
* 87010 87109: gap of 100 bp
* 87110 90855: contig of 3746 bp in length
* 90856 90955: gap of 100 bp
* 90956 96520: contig of 5565 bp in length
* 96521 96620: gap of 100 bp
* 96621 102321: contig of 5701 bp in length
* 102322 102421: gap of 100 bp
* 102422 108293: contig of 5872 bp in length
* 108294 108393: gap of 100 bp
* 108394 116689: contig of 8296 bp in length
* 116690 116789: gap of 100 bp
* 116790 125264: contig of 8475 bp in length
* 125265 125364: gap of 100 bp
* 125365 136354: contig of 10990 bp in length
* 136355 136454: gap of 100 bp
* 136455 145833: contig of 9379 bp in length.
Location/Qualifiers
1. 145833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20"
/clone="RP11-712N14"
/clone_lib="RPC1-11 Human Male BAC"
1. 1214
/note="assembly_fragment"
1315. 2673
/note="assembly_fragment"
2774. 4520
/note="assembly_fragment"
4621. 5961
/note="assembly_fragment"
6062. 7719
/note="assembly_fragment"
7820. 9799
/note="assembly_fragment"
9900. 11434
/note="assembly_fragment"
11535. 14382
/note="assembly_fragment"
14483. 17148
/note="assembly_fragment"
17249. 19239
/note="assembly_fragment"
19340. 21102
/note="assembly_fragment"
21203. 23371
/note="assembly_fragment"
23472. 25782
/note="assembly_fragment"
25883. 28824
/note="assembly_fragment"
28925. 31619
/note="assembly_fragment"
31720. 34679
/note="assembly_fragment"
34780. 37527
/note="assembly_fragment"
37628. 40146
/note="assembly_fragment"
40247. 43743
/note="assembly_fragment"

```

Matches 918; Conservative 0; Mismatches 0; Indels 1; Gaps 1,

Oy	1076	AGACGTCCTCCCTGTTTCTGAACTGCGACAGGGGAGAGACTCTTTCTCACTAGAGGCT	1135
Db	60571	ACAGCTCCCTCCTGTTTCTGAACTGCGACAGGGGAGAGACTCTTTCTCACTAGAGGCTC	60512
Oy	1136	TCCGGGAGTCCCTCAAGCTTCTACATCACTAGCCTGAATGACAGAGCTGTCTCTTGGATGATG	1195
Db	60511	TCCGGGAGTCCCTCAAGCTTCTACATCACTAGCCTGAATGACAGAGCTGTCTCTTGGATGATG	60452
Oy	1196	CCTAGGCCCAAGAGAGAGCCAAAGAGAGAAACAAGCTGTGACACCTAGAACCCCAATTC	1255
Db	60451	CCTAGGCCCAAGAGAGAGCCCAAGAGAGAAACAAGCTGTGACACCTAGAACCCCAATTC	60392
Oy	1256	AGCCTCTGGGGACCCCAAGAGAGAAAGGCTGTGACTCAAGGAGAGGAGGGTGGGACACAGA	1315
Db	60391	AGCCTCTGGGGACCCCAAGAGAGAAAGGCTGTGACTCAAGGAGAGAGGGTGGGACACAGA	60332
Oy	1316	GGTGCATCTAGGGTCCCACTGTACCTCTGTCTTTCTCTCTTAAGCCCTTAAAGATCAC	1375
Db	60331	GGTGCATCTAGGGTCCCACTGTACCTCTGTCTTTCTCTCTTAAGCCCTTAAAGATCAC	60272
Oy	1376	CTACTTCCTTCCAGTGGCCATGATGCCACCTGCGACCTTGTAGTGCAGATCAGAGAAAGTG	1435
Db	60271	CTACTTCCTTCCAGTGGCCATGATGCCACCTGCGACCTTGTGTAGTGCAGAGAAAGTG	60212
Oy	1436	GGACCAAGGGCGAGGGTCTCAAAAAGAAATAGAGCTCTGGGGGGTCTGACTTACTTAACT	1495
Db	60211	GGACCAAGGGCGAGGGTCTCAAAAAGAAATAGAGCTCTGGGGGGTCTGACTTACTTAACT	60152
Oy	1496	TCTTGAATTTGGGGTTTCAGTACCACTGTGATGTCGCTGTGTAGAGCCCATTTCTACA	1555
Db	60151	TCTTGAATTTGGGGTTTCAGTACCACTGTGATGTCGCTGTGTAGAGCCCATTTCTACA	60092
Oy	1556	TCCCCACCAATTACCAAGGCCCAACCCACAAGGTAGAAACAACCCCTAGAGTCAACGAGAA	1615
Db	60091	TCCCCACCAATTACCAAGGCCCAACCCACAAGGTAGAAACAACCCCTAGAGTCAACGAGAA	60032
Oy	1616	AGTCATTTTCAAAAATCTACAGTCTGCTTGAACCAACACATACCTCAAGAGGTAGG	1675
Db	60031	AGTCATTTTCAAAAATCTACAGTCTGCTTGAACCAACACATACCTCAAGAGGTAGG	59972
Oy	1676	ACTGTGGCTTAGAGGAAAGAAAGAAAGCTGAATGATCTTTACCCGAGACAGATCTTG	1735
Db	59971	ACTGTGGCTTAGAGGAAAGAAAGAAAGCTGAATGATCTTTACCCGAGACAGATCTTG	59912
Oy	1736	GATGTGTCAGAGCTGATGTGACCTCAAGACCAAGAAAGAACTTGGAGACGTGATGGTC	1795
Db	59911	GATGTGTCAGAGCTGATGTGACCTCAAGACCAAGAAAGAAAGCTTGGAGACGTGATGGTC	59852
Oy	1796	CTCAAAATGCCCCCATTTGAAGAACAGCCCACTCTTTTCTTTTGTGAGACGA	1855
Db	59851	CTCAAAATGCCCCCATTTGAAGAACAGCCCACTCTTTTCTTTTGTGAGACGA	59793
Oy	1856	GTTCTGACCTGTGCTCCAGTCTGAGATGCAATGACAGATCTCACTCACTGCAACCTCC	1915
Db	59792	GTTCTGACCTGTGCTCCAGTCTGAGATGCAATGACAGATCTCACTCACTGCAACCTCC	59733
Oy	1916	ATCTCCTGGAATTCAAATCTTCCGGCTAGGCTCCGAATTACTGCGGATTCAGAGCG	1975

RESULT	7
AF290985	
LOCUS	AF290985 786 bp mRNA linear PRI 22-APR-2002
DEFINITION	Homo sapiens Src-like adaptor protein-2 mRNA, complete cds.
ACCESSION	AF290985
VERSION	AF290985.1 GI:17351920
KEYWORDS	

ORGANISM *Homo sapiens*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini, Hominoidea; Homo.
REFERENCE
1 (Bases 1 to 786)
AUTHORS
Lozano, M.P. and McClellan, C.J.
TITLE
Submitted (22-JUL-2000) Brain Tumour Research Centre, Hospital for
Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada
JOURNAL
Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/tissue_type="thymus"
1..786
CDS

```

```

/codon_start=1
/product="Src-like adaptor protein-2"
/protein_id="AAL38197.1"
/db_xref="GI:17351921"
/translation="MGLSPFRKRSLLPSPLSSSVQGGPVTWEAKRSKATAVALGSP
AGGAELSLRLGEPLTIVSDGDGMVTAVLSEVSGEYNIPEYVAATVSHGVLVEGLSR
KAEELILPONGGAFVFLRSQTRGCSYSLVRSIRASMDIRIRYRIHCLNDGMYLY
SPRTPLFSEALAVDHYSELADIDICCLKEPCVLDORAPLPGAGDIPLPVTQRIPLNWM
ELDLSPLFEATGSESLSEGRCRSLFTSLNDEAVSLDA"

```

Query Match	30.6%	Score 786	DB 9	Length 786
Best Local Similarity	100.0%	Fred. No. 0		
Matches 786	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy 415	ATGGAGATCTCCCGACAGAGAAAATCTGTGCAAGCCCAAGCTTGAGTTCCTCTGTC	474		
Db 1	ATGGAGATCTCCCGACAGAGAAAATCTGTGCAAGCCCAAGCTTGAGTTCCTCTGTC	60		
Qy 475	CAAGGCGACAGGACCTGTGACCAATGAGACAAAGAGAACCAAGGCCAAGCCGTGGCCCTG	534		
Db 61	CAAGGCGACAGGACCTGTGACCAATGAGACAAAGAGAGAACCAAGGCCAAGCCGTGGCCCTG	120		
Qy 535	GGCAGTTTCGGGCAAGTGGCCGCGGCGCAGCTGTGCTGAGACTCGGAGAGCCATTGACC	594		
Db 121	GGCAGTTTCGGGCAAGTGGCCGCGGCGCAGCTGTGCTGAGACTCGGAGAGCCATTGACC	180		
Qy 595	ATGCTCTCTGAAGATGAGACATGTGTGACAGCGTGTCTGAAATGCTTACGACAGAGAGAT	654		
Db 181	ATGCTCTCTGAAGATGAGACATGTGTGACAGCGTGTCTGAAATGCTTACGACAGAGAGAT	240		
Qy 655	AACATCCCGAGGTCACATGTGGCAAAATCTCCCATGGATGGATGAGTATGAGGCGCTGAC	714		
Db 241	AACATCCCGAGGTCACATGTGGCAAAATCTCCCATGGATGGATGAGTATGAGGCGCTGAC	300		
Qy 715	AGGAGAGAAAGCAGAGGAACTGCTGTTGTAACCTGGAGAACCTTGGAGGGGCTTTCTCATC	774		
Db 301	AGGAGAGAAAGCAGAGGAACTGCTGTTGTAACCTGGAGAACCTTGGAGGGGCTTTCTCATC	360		
Qy 775	CGGAGAGCCAGACCAAGAGAGGCTTTACTCTCTGACGTCCGCTCAACCGGCTTCA	834		

```

Db 361 CCGGAGACCGACGACGAGAGAGGCTTACTCTCTGCACTCCGCTACGCCGCTCGCA 420
Oy 835 TCCTGGGACCGGATCAGACACTACAGATCACTGCTTGAACATGCTGGCTGTACATC 894
Db 421 TCCTGGGACCGGATCAGACACTACAGATCACTGCTTGAACATGCTGGCTGTACATC 480
Oy 895 TCACCGGCGCTCACCCTTCCCTCACCAGGCGCTGGTGAACCATTTACTTGAAGCTGGG 954
Db 481 TCACCGGCGCTCACCCTTCCCTCACCAGGCGCTGGTGAACCATTTACTTGAAGCTGGG 540
Oy 955 GATGACATCTGCTGCTTACTCAAGAGCCCTGTGTCTTGAAGAGGCTGGCCCTCCT 1014
Db 541 GATGACATCTGCTGCTTACTCAAGAGCCCTGTGTCTTGAAGAGGCTGGCCCTCCT 600
Oy 1015 GGCAGAGATATACCCCTACCTGCTGACAGAGAGACACACTCACTGGAAGAAGCTG 1074
Db 601 GGCAGAGATATACCCCTACCTGCTGACAGAGAGACACACTCACTGGAAGAAGCTG 660
Oy 1075 GACAGCTCCCTCTGCTTTTGTGAAGCTGCACAGAGGAGAGTCTTCTCACTGAAGGT 1134
Db 661 GACAGCTCCCTCTGCTTTTGTGAAGCTGCACAGAGGAGAGTCTTCTCACTGAAGGT 720
Oy 1135 CTCGGGAGATCCCTCAGCTTCTACATCAGCCCTGAATGACAGAGGCTGTCTTTGGATGAT 1194
Db 721 CTCGGGAGATCCCTCAGCTTCTACATCAGCCCTGAATGACAGAGGCTGTCTTTGGATGAT 780
Oy 1195 GCCTAG 1200
Db 781 GCCTAG 786

RESULT 8
AF326353 786 bp mRNA linear PRI 08-NOV-2001
LOCUS AF326353
DEFINITION Homo sapiens Src-like adaptor protein-2 mRNA, complete cds.
ACCESSION AF326353.1 GI:16797891
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 786)
    Holland, S.J., Liao, X.C., Mendenhall, M.K., Zhou, X., Pardo, J.,
    Chu, P., Spencer, C., Fu, A.C., Sheng, N., Yu, P., Pali, E., Nagin, A.,
    Shen, M., Yu, S., Chan, E., Wu, X., Li, C., Moissetschlager, M.,
    Aversa, G., Kolbinger, F., Bennett, M.K., Moliniaux, S., Luo, Y.,
    Pavan, D.G., Mancebo, H.S.Y. and Wu, J.
    Functional Cloning of Src-like Adaptor Protein-2 (SLAP-2), a Novel
    Inhibitor of Antigen Receptor Signaling
    J. Exp. Med. 194 (9), 1263-1276 (2001)
JOURNAL
MEDLINE 2153259
PUBMED 11696592
TITLE 2 (bases 1 to 786)
AUTHORS Holland, S.J., Mendenhall, M.K., Zhou, X., Spencer, C., Pardo, J.,
        Fu, A.C., Sheng, N., Shen, M., Liao, X.C., Luo, Y., Pavan, D.G.,
        Mancebo, H.S.Y. and Wu, J.
FEATURES
    source
        1..786
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="20"
            1..786
            /note="SLAP2, SH2 and SH3 domain-containing adaptor"
            /codon_start=1
            /product="Src-like adaptor protein-2"
            /protein_id="AAI29204.1"
            /db_xref="GI:16797892"
            /translation="MGSLSRRLKSLPSPSLSSVSGQGPVTMEASRKAVALGSPF
            AGPAELRLTGEPLTIVSEDDWMTVLSEVSGREINIPSVHAKVSHMLYEGLSRE

```

```

BASE COUNT      162 a      234 c      231 g      159 t
ORIGIN
Query Match      30.6%; Score 786; DB 9; Length 786;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
KAEELLLPQNGGAFILRESOTRGRGYSLSVLSPPASMDRIHRYHICLDNGWLYI
SPRLTPPSLQALVDHYSLIADDICLLKEPCVLRAGPLPKGDIPLPVTVQRPPLMWK
ELDSSLSEHATGESLSLSBGLSESLFSLISLNDVAVSLDDA"
Oy 415 ATGGGAGATCTGCCCGACGAGAAAGAAATCTGCGCAAGCCCAAGCTTGAATTCCTGTC 474
Db 1 ATGGGAGATCTGCCCGACGAGAAAGAAATCTGCGCAAGCCCAAGCTTGAATTCCTGTC 60
Oy 475 CAGGCGCAGGACCTGTGACCATGGAAGCAGAGAGAGCAGAGCCAGCCGTGGCCCTG 534
Db 61 CAGGCGCAGGACCTGTGACCATGGAAGCAGAGAGAGCAGAGCCAGCCGTGGCCCTG 120
Oy 535 GGCATTTTCCCGGACAGTGGCCCCGCGAGCTTGCCTGAGACTGGGGAGCCATTGACC 594
Db 121 GGCATTTTCCCGGACAGTGGCCCCGCGAGCTTGCCTGAGACTGGGGAGCCATTGACC 180
Oy 595 ATCTCTCTGAGAGATGAGAGATGTTGAGCGATGCTGTCTGAAGTCTCAGGCAGAGATAT 654
Db 181 ATCTCTCTGAGAGATGAGAGATGTTGAGCGATGCTGTCTGAAGTCTCAGGCAGAGATAT 240
Oy 655 AACATCCCCACAGCTTCCACGTGGCCAAAGTCTCCCATGGGTGCTGTATGAGAGCCCTGAGC 714
Db 241 AACATCCCCACAGCTTCCACGTGGCCAAAGTCTCCCATGGGTGCTGTATGAGAGCCCTGAGC 300
Oy 715 AGGAGAAAGCAGAGAGAACTGCTGTTTGAATCTGGGAAACCTCGAGAGGCTTCTCATC 774
Db 301 AGGAGAAAGCAGAGAGAACTGCTGTTTGAATCTGGGAAACCTCGAGAGGCTTCTCATC 360
Oy 775 CCGGAGAGCAGACACAGAGAGAGGCTTACTCTGTCTGTCAGTCCGCTCAGGCCCTTGCA 834
Db 361 CCGGAGAGCAGACACAGAGAGAGGCTTACTCTGTCTGTCAGTCCGCTCAGGCCCTTGCA 420
Oy 835 TCCTGGGACCGGATCAGACACTACAGATCACTGCTTGAACATGCTGGCTGTACATC 894
Db 421 TCCTGGGACCGGATCAGACACTACAGATCACTGCTTGAACATGCTGGCTGTACATC 480
Oy 895 TCACCGGCGCTCACCCTTCCCTCACCAGGCGCTGGTGAACCATTTACTTGAAGCTGGG 954
Db 481 TCACCGGCGCTCACCCTTCCCTCACCAGGCGCTGGTGAACCATTTACTTGAAGCTGGG 540
Oy 955 GATGACATCTGCTGCTTACTCAAGAGCCCTGTGTCTTGAAGAGGCTGGCCCTCCT 1014
Db 541 GATGACATCTGCTGCTTACTCAAGAGCCCTGTGTCTTGAAGAGGCTGGCCCTCCT 600
Oy 1015 GGCAGAGATATACCCCTACCTGCTGACAGAGAGACACACTCACTGGAAGAAGCTG 1074
Db 601 GGCAGAGATATACCCCTACCTGCTGACAGAGAGACACACTCACTGGAAGAAGCTG 660
Oy 1075 GACAGCTCCCTCTGCTTTTGTGAAGCTGCACAGAGGAGAGTCTTCTCACTGAAGGT 1134
Db 661 GACAGCTCCCTCTGCTTTTGTGAAGCTGCACAGAGGAGAGTCTTCTCACTGAAGGT 720
Oy 1135 CTCGGGAGATCCCTCAGCTTCTACATCAGCCCTGAATGACAGAGGCTGTCTTTGGATGAT 1194
Db 721 CTCGGGAGATCCCTCAGCTTCTACATCAGCCCTGAATGACAGAGGCTGTCTTTGGATGAT 780
Oy 1195 GCCTAG 1200
Db 781 GCCTAG 786

RESULT 9
AF290986 737 bp mRNA linear PRI 22-APR-2002
LOCUS AF290986
DEFINITION Homo sapiens Src-like adaptor protein-2 splice isoform mRNA,
ACCESSION AF290986

```

VERSION	AF290986.1	GI:17351922
KEYWORDS		
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 737)	
AUTHORS	Loreto,M.P., and McLade,C.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (28-JUL-2000) Brain Tumour Research Centre, Hospital for Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada	
FEATURES	location/Qualifiers	
source	1..737	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="20"	
	/tissue_type="thymus"	
CDS	1..633	
	/note="SLAP-2-v"	
	/codon_start=1	
	/product="Src-like adaptor protein-2 splice isoform"	
	/protein_id="ABJ8198.1"	
	/db_xref="GI:17351923"	
	/translation="MQSLPSRRKSLPSLSISSVQGGPVTMEARSKATAVALGSFPL AGGAELSLRLGEPLTIVSEDDMMWTLSEVGREYNIPSVYAKVSHQMLVEGLSR KAELLILPNPGCAFLIRESDTRGSYSISLRSPASMPRI.RHRIHICLNMCLYI PRTPFSLLALVDHYSEGMAPMGQTPTPCACADTTQLERAGLPVPF"	
BASE COUNT	152 a 219 c 218 g 148 t	
ORIGIN		
Query Match	20.8%; Score 534; DB 9; Length 737;	
Best Local Similarity	100.0%; Pred. No. 1.9e-291;	
Matches 534; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Oy	415 ATGGAGAATCTCCCGCAGAGAGAAATCTGCCAAGCCCAAGCTTAGTGTCCTGTGC	474
Dd	1 ATGGAGAATCTCCCGCAGAGAGAAATCTGCCAAGCCCAAGCTTAGTGTCCTGTGC	60
Oy	475 CAAGGCCAGGACCTGTGACCATGAGACAGAGAGAGACAAGCCACAGCCGTGGCCCTG	534
Dd	61 CAAGGCCAGGACCTGTGACCATGAGAGAGAGAGACAAGCCACAGCCGTGGCCCTG	120
Oy	535 GGCAGTTTCCCGCAGGTGGCCGGCCGAGACTGTGCTGACGACTCGGGGAGCCAATTGACC	594
Dd	121 GGCAGTTTCCCGCAGGTGGCCGGCCGAGACTGTGCTGACGACTCGGGGAGCCAATTGACC	180
Oy	595 ATGCTCTTGAAGATGACGACTGGTGAACCGTGTCTGTGAAGTCTCAGGACAGAGATAT	654
Dd	181 ATGCTCTTGAAGATGACGACTGGTGAACCGTGTCTGTGAAGTCTCAGGACAGAGATAT	240
Oy	655 AACATCCCCAGGCTTCACGTGGCCAAAGTCTCCATGGGTGGCTGTATAGGGGCTGAGC	714
Dd	241 AACATCCCCAGGCTTCACGTGGCCAAAGTCTCCATGGGTGGCTGTATAGGGGCTGAGC	300
Oy	715 AGGGAAGAAGCAGAGAAATGCTGTGTACTCTGGGAACCTTGGAGGGGCTTCTTCATC	774
Dd	301 AGGGAAGAAGCAGAGAAATGCTGTGTACTCTGGGAACCTTGGAGGGGCTTCTTCATC	360
Oy	775 CGGGAGAGCCAGACCGAGAGAGGCTTTATCTCTGTTCAGTCTGCCTCAGCCGACCTGCA	834
Dd	361 CGGGAGAGCCAGACCGAGAGAGGCTTTATCTCTGTTCAGTCTGCCTCAGCCGACCTGCA	420
Oy	835 TCTTGGGACCGGATCAGACACTACAGATCCACGTGCTTGCACAAATGGCTGCTTAATC	894
Dd	421 TCTTGGGACCGGATCAGACACTACAGATCCACGTGCTTGCACAAATGGCTGCTTAATC	480
Oy	895 TCACGGGCTCACCTTCCCTCAGTCCACGAGCCCTGTGGACCATTAATTCTGAG	948
Dd	481 TCACGGGCTCACCTTCCCTCAGTCCACGAGCCCTGTGGACCATTAATTCTGAG	534

LOCUS	HS460J8	66741 bp	DNA	linear	PR1 23-JUL-2001							
DEFINITION	Human DNA sequence from clone RP3-460J8 on chromosome 20q11.21-11.23 Contains the 3' end of the gene for a novel protein similar to N-myc downstream regulated (NDRG1), the 5' end of a gene encoding a novel protein tyrosine kinase, ESTs, STS and GSSs, complete sequence.											
ACCESSION	AL031662											
VERSION	AL031662.26	GI:97169301										
KEYWORDS	HTG; NDRG1; SH2 domain.											
SOURCE	human.											
ORGANISM	Homo sapiens											
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eumetaria; Primates; Catarrhini; Homnidae; Homo.											
AUTHORS	1 (bases 1 to 66741)											
TITLE	Direct Submission											
JOURNAL	Submitted (23-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgery@sanger.ac.uk											
COMMENT	On Aug 7, 2000 this sequence version replaced gi:6425549. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences without a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPeP; Information on the WormPeP database can be found at http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20 IMPORTANT: This sequence is not the entire insert of clone RP3-460J8. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP3-460J8 is at 66741 in this sequence. The true left end of clone RP3-460J8 is at 41767 in this sequence. The true right end of clone RP5-977B1 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP3-460J8 is from the library RPCI-3 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pCYPAC2.											
FEATURES	Location/Qualifiers											
SOURCE	1..66741											
	/organism="Homo sapiens"											
	/db_xref="taxon:9606"											
	/chromosome="20"											
	/map="q11.21-11.23"											
	/clone="RP3-460J8"											
	/clone_1lb="RPCI-3"											
	complement(50..544)											
	/note="match: GSS: Em:B45150"											
gene	complement(240..7128)											
	/gene="d9j77B1.1"											
mrna	complement(join(240..339,6995..7128))											
	/gene="d9j77B1.1"											
	/product="d460J8.2 (novel protein tyrosine kinase with Src homology 2 (SH2) domain)"											
	Src homology 2 (SH2) domain"											
	/note="match: cDNAs: Em:BK025645											
	match: ESTs: Em:BG178487											
	/evidence="not_experimental"											
	complement(join(240..339,6995..7085))											
	/gene="d9j77B1.1"											
CDS												

```
/note="Continues in Em:AL050318 as dJ977B1.2"
/codon_start=1
/evidence=not experimental
/product="dJ460J8.2 (novel protein tyrosine kinase with
src homology 2 (SH2) domain )"
/protein_id="CAC4645.1"
/db_xref="GI:15020830"
/translation="MGSLSRKRSLPSPSLSSVQGGPVTWEAKRATKVALGSFP
AGSPALSLRLGEPLTIVS"
repeat_region
536..565
/note="15 copies 2 mer ag 93% conserved"
2375..2402
/note="14 copies 2 mer ta 100% conserved"
2783..3155
/note="match: GSS: Em:AQ0807191"
complement(6983..7482)
/note="match: GSS: Em:AQ056467"
complement(6985..7492)
/note="match: GSS: Em:AQ056478"
7486..7925
/note="match: STS: Em:HS427J1S"
10615..10640
/note="13 copies 2 mer tc 92% conserved"
13593..13961
/note="MER21B repeat: matches 422..785 of consensus"
14248..14297
/note="MER21B repeat: matches 374..422 of consensus"
14594..14614
/note="MER21B repeat: matches 355..374 of consensus"
14746..15078
/note="MER21B repeat: matches 1..355 of consensus"
complement(17516)
17519..17686
/note="match: GSS: Em:AQ067993"
complement(17521)
complement(17522)
polyA_site
polyA_site
gene
MRNA
/gene="dJ469A13.3"
complement(17523..54534)
22247..22282,26087..26134,30796..30847,30986..31051,
32039..32142,37091..37147,46566..46652,48268..48328,
50163..50225,53242..53362,54429..54534)
/gene="dJ469A13.3"
/product="dJ460J8.1 (continued from dJ469A13.3 in
Em:AL132768)"
/note="match: CDNAS: Em:AB033922 Em:U52073 Em:DB7953
Em:X92845 Em:AF004162 Em:U60593 Em:AF045564 Em:AB033921
Em:AF159092 Em:AF147402 Em:W59814
match: ESTs: Em:AA178726 Em:AA039000 Em:AM003952 Em:W89263
Em:AA325826 Em:AI230982 Em:AA162360 Em:AA445016
Em:AV002395 Em:T85147 Em:AI786673 Em:AA113437 Em:AI004026
Em:T88705 Em:AI786673 Em:AI181197 Em:AA764653 Em:AA316771
Em:AU035165 Em:AV002368"
/evidence=not experimental
complement(17542..17547)
/gene="dJ469A13.3"
17555..17811
/note="match: STS: Em:G19945"
join(19202..19451,22110..22282,26087..26127,30796..30847)
/note="match: STS: Em:G23762"
complement(join(19202..19451,20551..20589,22110..22161,
22247..22282,26087..26134,30796..30847,30986..31051,
32039..32142,37091..37147,46566..46652,48268..48328,
50163..50225,53242..53362,54429..54534))
/gene="dJ469A13.3"
/note="novel protein (FLJ13556) similar to N-myc
demonstrated regulated (NDRG1)
match: proteins: Sw:Q62433 Sw:Q92597 Tr:Q9Z219 Sw:P97862"
/codon_start=1
/evidence=not experimental
/product="dJ460J8.1 (continued from dJ469A13.3 in
Em:AL132768)"
/protein_id="CAB56525.1"
```

```
/db_xref="GI:6687781"
/db_xref="SWISS-PROT:Q9UGV2"
/translation="EHDIETTHGVAVHITGLPKGNRPVILTYHDIGLNKSCFNAAF
NEDMOETIOTHFAVCHVAPGGGAPSPPTGYOYPTMDELAEMLPVLTLSLSII
GIGGAGAGVILSRFALNHPELVBEGLVLIINTDPAKQMTDMASLSGLTTNVVDITLA
HRRGFEQLANDLIDTQYRMHIAODINODLDPLNKSNGRRDIEIRPIILGQNDKS
KTLKSTLLVGVNDSFAVAAYVECSRNLNINTTLKADCGGLPVVQGLTEAFK
YFLQGGIPIYVQLSHLSTVESPSASWTRLSRTRTSSLSGSPFSRSVTNOS
DQDQSCSPVLDLRHQTWEVSC"
complement(24246..24780)
/gene="dJ469A13.3"
complement(22980..25556)
/note="match: GSS: Em:AQ0592789"
complement(27980..28556)
/gene="dJ469A13.3"
/note="match: GSS: Em:AQ08867"
34313..34420
/note="HY1 repeat: matches 1..109 of consensus"
34505..34560
/note="28 copies 2 mer ta 78% conserved"
34564..34611
/note="24 copies 2 mer ta 79% conserved"
42553..43658
/note="MER52C repeat: matches 1..1278 of consensus"
47989..48470
/note="match: GSS: Em:AQ0672417"
49391..49480
/note="45 copies 2 mer ta 76% conserved"
52922..53120
/note="match: STS: Em:G04621"
complement(53426..54032)
/gene="dJ469A13.3"
/note="match: GSS: Em:AQ314824
match: STS: Em:G55853"
54054..54477
/note="match: GSS: Em:AQ310681"
57509..57838
/note="match: STS: Em:G07504"
complement(57696..58013)
/note="match: STS: Em:G07632 Em:G07634"
58856..59001
/note="73 copies 2 mer at 83% conserved"
complement(58930..59281)
/note="match: GSS: Em:AQ067563"
59062..59236
/note="TIGER2 repeat: matches 2541..2714 of consensus"
59738..60495
/note="TIGER2 repeat: matches 1780..2541 of consensus"
60719..61831
/note="MER11C repeat: matches 1..1071 of consensus"
62547..63174
/note="TIGER2 repeat: matches 1165..1784 of consensus"
63475..63703
/note="TIGER2 repeat: matches 125..354 of consensus"
64017..64063
Query Match 14.0%; Score 359; DB 9; Length 66741;
Best Local Similarity 100.0%; Pred.No. 3e-192;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 GTGCGAGCTAGAGCTCCCAAGGACCCACGCGTGTCTCTGTGACAGAGCTCAAGGGCC 72
DB 11945 GTGCGAGCTAGAGCTCCCAAGGACCCACGCGTGTCTCTGTGACAGAGCTCAAGGGCC 11886
QY 73 CTGGGCTTCCCTCCCTGCTGGCTGCTGCTGGAGAGGTTCCCAAGTCAAAATCCCTTA 132
DB 11885 CTGGGCTTCCCTCCCTGCTGGCTGCTGCTGGAGAGGTTCCCAAGTCAAAATCCCTTA 11826
QY 133 AGAGCATGTGGGAGAGTATCATCTCTGTGTGACAAATCTGCTAGTCAGACAGATGCT 192
DB 11825 AGAGCATGTGGGAGAGTATCATCTCTGTGTGACAAATCTGCTAGTCAGACAGATGCT 11766
QY 199 GAGCTACCAACCAACCACTAGCGCTCTCCGAGAGCTTCCCAAGCTGAGAGCTTCT 252
|||||
```

Db 11765 GAGCTACCAACCAACCACTAGCCTCTCTCCGAAGATCTCCAGAGCTGAGAGATTTCT 11706
 OY 253 GGGTGTCTAGGAGACCAAGACACTGGCAGACTTCCAGAGAGGCCCCCAAGAGCTTAACT 312
 Db 11705 GGGTGTCTAGGAGACCAAGACACTGGCAGACTTCCAGAGAGGCCCCCAAGAGCTTAACT 11646
 OY 313 GTCCAGCAGACAGATGCTCTCAGCAGAGCTCTCTCCAGAGCTTGTATGACAAACCA 371
 Db 11645 GTCCAGCAGACAGATGCTCTCAGCAGAGCTCTCTCCAGAGCTTGTATGACAAACCA 11587

RESULT 11

AC026539 145833 bp DNA linear HTG 27-APR-2000
 LOCUS Homo sapiens chromosome 20 clone RP11-712N14 map 20, WORKING DRAFT
 DEFINITION SEQUENCE, 37 unordered pieces.
 AC026539
 AC026539.2 GI:7656813
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 20, clone RP11-712N14
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 145833)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Baetien, V., Bedalov, F.,
 Boguslavsky, L., Bouckwalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Chao, P., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeRubeis, K., Dewar, K., Diaz, J., S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lebecky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGurt, A., McKernan, K., McPherson, R.,
 Meldrum, J., Meneses, L., Mihova, T., Miranda, C., Mlangeni, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T., M., Oliver, J., Peterson, K., Pierre, N.,
 Pisan, C., Pollard, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tittell, A., Travers, M., Trigg, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT
 Direct Submission
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 27, 2000 this sequence version replaced g1.7283243.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L7115

Center clone name: 712.N.14

Sequencing vector: M13, M77815, 100% of reads

Chemistry: Dye-terminator Big Dye, 100% of reads

Assembly program: Phrap, version 0.960731

Consensus quality: 125577 bases at least Q40

Consensus quality: 135703 bases at least Q30

Consensus quality: 135933 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 142233; sum-of-contigs

Quality coverage: 2.6 in Q20 bases; agarose-fp
 Quality coverage: 3.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 37 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1214: contig of 1214 bp in length
 * 1215 1314: gap of 100 bp
 * 1315 2673: contig of 1359 bp in length
 * 2674 2773: gap of 100 bp
 * 2774 4520: contig of 1747 bp in length
 * 4521 4620: gap of 100 bp
 * 4621 5961: contig of 1341 bp in length
 * 5962 6061: gap of 100 bp
 * 6062 7719: contig of 1658 bp in length
 * 7720 7819: gap of 100 bp
 * 7820 9799: contig of 1980 bp in length
 * 9800 9899: gap of 100 bp
 * 9900 11434: contig of 1535 bp in length
 * 11435 11534: gap of 100 bp
 * 11535 14382: contig of 2848 bp in length
 * 14383 14482: gap of 100 bp
 * 14483 17148: contig of 2666 bp in length
 * 17149 17248: gap of 100 bp
 * 17249 19239: contig of 1991 bp in length
 * 19240 19339: gap of 100 bp
 * 19340 21102: contig of 1763 bp in length
 * 21103 21202: gap of 100 bp
 * 21203 23371: contig of 2169 bp in length
 * 23372 23471: gap of 100 bp
 * 23472 25782: contig of 2311 bp in length
 * 25783 25882: gap of 100 bp
 * 25883 28824: contig of 2942 bp in length
 * 28825 28924: gap of 100 bp
 * 28925 31619: contig of 2695 bp in length
 * 31620 31719: gap of 100 bp
 * 31720 34675: contig of 2960 bp in length
 * 34680 34779: gap of 100 bp
 * 34780 37527: contig of 2748 bp in length
 * 37528 37627: gap of 100 bp
 * 37628 40146: contig of 2519 bp in length
 * 40147 40246: gap of 100 bp
 * 40247 43743: contig of 3497 bp in length
 * 43744 43843: gap of 100 bp
 * 43844 47133: contig of 3290 bp in length
 * 47134 47233: gap of 100 bp
 * 47234 51023: contig of 3790 bp in length
 * 51024 51123: gap of 100 bp
 * 51124 54935: contig of 3812 bp in length
 * 54936 55035: gap of 100 bp
 * 55036 59553: contig of 4518 bp in length
 * 59554 59653: gap of 100 bp
 * 59654 64428: contig of 4775 bp in length
 * 64429 64528: gap of 100 bp
 * 64529 69211: contig of 4683 bp in length
 * 69212 69311: gap of 100 bp
 * 69312 72901: contig of 3590 bp in length
 * 72902 73001: gap of 100 bp
 * 73002 76724: contig of 3723 bp in length
 * 76725 76824: gap of 100 bp
 * 76825 81179: contig of 4355 bp in length
 * 81180 81279: gap of 100 bp
 * 81280 87009: contig of 5730 bp in length
 * 87010 87109: gap of 100 bp
 * 87110 90855: contig of 3746 bp in length
 * 90856 90955: gap of 100 bp
 * 90956 96520: contig of 5565 bp in length
 * 96521 96620: gap of 100 bp

FEATURES
source

```

/organism:"Homo sapiens"
/db xref:"Taxon:9606"
/chromosome:"20"
/map="20"
/clone:"RP11-712N14"
/clone_lib:"RP11 Human Male BAC"
1..1214
misc_feature
  /note="assembly_fragment"
  115..2673
misc_feature
  /note="assembly_fragment"
  2774..4520
misc_feature
  /note="assembly_fragment"
  4621..5961
misc_feature
  /note="assembly_fragment"
  6062..7719
misc_feature
  /note="assembly_fragment"
  7820..9799
misc_feature
  /note="assembly_fragment"
  9900..11434
misc_feature
  /note="assembly_fragment"
  11535..14382
misc_feature
  /note="assembly_fragment"
  14483..17148
misc_feature
  /note="assembly_fragment"
  17249..19239
misc_feature
  /note="assembly_fragment"
  19340..21102
misc_feature
  /note="assembly_fragment"
  21203..23171
misc_feature
  /note="assembly_fragment"
  23472..25882
misc_feature
  /note="assembly_fragment"
  25883..28824
misc_feature
  /note="assembly_fragment"
  28925..31619
misc_feature
  /note="assembly_fragment"
  31720..34679
misc_feature
  /note="assembly_fragment"
  34780..37527
misc_feature
  /note="assembly_fragment"
  37628..40146
misc_feature
  /note="assembly_fragment"
  40247..43743
misc_feature
  /note="assembly_fragment"
  43844..47133
misc_feature
  /note="assembly_fragment"
  47234..51023
misc_feature
  /note="assembly_fragment"
  51124..54935
misc_feature
  /note="assembly_fragment"
  clone_end:17
vector_side:right"
55036..59553
/note="assembly_fragment"

```

Query Match	14.0%	Score 359;	DB 2;	Length 145833;
Best Local Similarity	100.0%;	Pred. No. 2,7e+192;		
Matches 359;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
07	13	GTGGAGCTGAGAGCTCAAGACCCACCCCTGTGTCTTGTGACAGACTCAAAAGGCC	72	

RESULT 12	
AL353754/c	
LOCUS	17316 bp DNA linear PRI 30-JAN-2002
DEFINITION	Human DNA sequence from clone RP1-401E9 on chromosome 10, complete sequence.
ACCESSION	AL353754
VERSION	AL353754.10 GI:18476549
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 173316)
Griffiths, C.
Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,

COMMENT

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Swi, SWISSPROT; Tr1, TREMBL; Wp1, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Map/Ch10> RP11-401E9 is from the library RP11-2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/backpac/home.htm>

VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-401E9 The true left end of clone RP11-554F11 is at 118361 in this sequence. The true right end of clone RP11-299D3 is at 102633 in this sequence.

Location/Qualifiers

source 1. 173316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-401B9"
/clone_1ib="RP11-1.2"
misc_feature complement(65129..65132)
/note="1189 bases of ISS (001735) removed here. This sequence represents the duplicated flanking sequence of the ISS."

BASE COUNT 48761 a 33467 c 35550 g 55538 t
ORIGIN

Query Match 2.9%; Score 74; DB 9; Length 173316;
Best Local Similarity 100.0%; Pred. No. 2.2e-30;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2035 TTGGCCAGCGCTGTGCAACTCTGACCTGATGATCCACCACCTTGAGCTCCCAA 2094
Db 109183 TTGGCCAGCGCTGTGCAACTCTGACCTGATGATCCACCACCTTGAGCTCCCAA 109124

OY 2095 GTGCTGGGATTACA 2108
Db 109123 GTGCTGGGATTACA 109110

RESULT 13
LOCUS AL157361 151603 bp DNA linear PRI 24-MAY-2000
DEFINITION Human DNA sequence from clone RP11-132u12 on chromosome 13q21.33-22.3 contains an STS and GSS, complete sequence.
ACCESSION AL157361
VERSION AL157361.6 GI:7899161
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Phyllome.B.
DIRECT SUBMISSION
Submitted (24-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On May 17, 2000 this sequence version replaced gi:7739119.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unseq' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information at the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-132u12 is from the library RP11-1.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR: pBACs.6
This sequence is the entire insert of clone RP11-132u12. The true left end of clone RP11-15942 is at 131986 in this sequence.

FEATURES
location/Qualifiers

source 1. 151603
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q21.33-22.3"
/clone="RP11-132u12"
/clone_1ib="RP11-1.1"
3. 79
repeat_region /note="MULTI repeat: matches 472..549 of consensus"
315..1757
repeat_region /note="L1PB3 repeat: matches 4707..6150 of consensus"
2063..2269
repeat_region /note="MLT2CB repeat: matches 1..222 of consensus"
2270..2666
repeat_region /note="MSTA repeat: matches 3..426 of consensus"
2667..2963
repeat_region /note="MLT2CB repeat: matches 222..466 of consensus"
3022..3073
repeat_region /note="MLT2CB repeat: matches 455..504 of consensus"
/note="complement(3961..4381)
/note="match: GSS: Em:AQ089797"
/note="complement(3964..4397)
/note="match: GSS: Em:AQ31302"
/note="complement(3990..4397)
/note="match: GSS: Em:AQ361074"
5414..5579
repeat_region /note="MERSA repeat: matches 5..187 of consensus"
7569..7652
repeat_region /note="MERSA repeat: matches 97..183 of consensus"
8606..8883
repeat_region /note="L1MBC repeat: matches 165..1545 of consensus"
9438..9551
repeat_region /note="L1M4A repeat: matches 2084..2197 of consensus"
9726..10188
repeat_region /note="L1M4A repeat: matches 2302..2774 of consensus"
10189..10484
repeat_region /note="L1M4A repeat: matches 1..296 of consensus"
10495..10798
repeat_region /note="L1M4A repeat: matches 1..303 of consensus"
10799..12407
repeat_region /note="L1M4A repeat: matches 2783..4427 of consensus"
12408..12713
repeat_region /note="L1M4A repeat: matches 1..308 of consensus"
12714..14632
repeat_region /note="L1M4A repeat: matches 4427..6254 of consensus"
14659..15748
repeat_region /note="L1M4 repeat: matches 2138..3260 of consensus"
15796..16104
repeat_region /note="L1M4 repeat: matches 1..309 of consensus"
16187..16492
repeat_region /note="L1M4 repeat: matches 3252..3570 of consensus"
16496..16927
repeat_region /note="L1M4 repeat: matches 3832..4256 of consensus"
16928..17175
repeat_region /note="L1M4 repeat: matches 39..289 of consensus"
17176..17258
repeat_region /note="L1M4 repeat: matches 4256..4339 of consensus"
17363..17472
repeat_region /note="L1M4 repeat: matches 4497..4606 of consensus"
17533..17916
repeat_region /note="L1M4 repeat: matches 1..365 of consensus"
17917..18332
repeat_region /note="L1M4 repeat: matches 25..442 of consensus"
18355..18420
repeat_region /note="MIR repeat: matches 145..211 of consensus"
18708..19305
repeat_region /note="L1PB1 repeat: matches 5542..6138 of consensus"
19768..19889
repeat_region /note="61 copies 2 mer at 644 conserved"
20002..20298
repeat_region /note="L1M1 repeat: matches 5534..5863 of consensus"
22697..22911
repeat_region /note="MERSA repeat: matches 9..217 of consensus"

```

repeat_region 23502..23547
/note="23 copies 2 mer at 78% conserved"
repeat_region 24310..24754
/note="L2 repeat: matches 2275..2749 of consensus"
repeat_region 25348..25561
/note="MIR repeat: matches 8..224 of consensus"
misc_feature 25847..26252
/note="match: GSS: Em:AQ602207"
repeat_region 25911..25955
/note="MIR repeat: matches 99..145 of consensus"
repeat_region 27540..27585
/note="MIR repeat: matches 216..261 of consensus"
repeat_region 28853..29156
/note="ALUX repeat: matches 1..309 of consensus"
repeat_region 31217..31248
/note="16 copies 2 mer ac 96% conserved"
repeat_region 31397..31608
/note="HAL1 repeat: matches 1191..1409 of consensus"
repeat_region 31670..31963
/note="ALUX repeat: matches 1..291 of consensus"
repeat_region 32072..32223
/note="MSTC repeat: matches 1..152 of consensus"
repeat_region 32315..32428
/note="MSTC repeat: matches 248..363 of consensus"
repeat_region 32429..32664
/note="LIMB7 repeat: matches 309..553 of consensus"
repeat_region 32668..34742
/note="LIMB7 repeat: matches 4075..6146 of consensus"
repeat_region 34775..35227
/note="LIMB7 repeat: matches 538..970 of consensus"
repeat_region 35228..35531
/note="LIMB7 repeat: matches 1..301 of consensus"
repeat_region 35532..36720
/note="LIMB7 repeat: matches 970..2189 of consensus"
repeat_region 36886..37021
/note="LIMB7 repeat: matches 2204..2341 of consensus"
repeat_region 37029..37220
/note="LIMB7 repeat: matches 107..304 of consensus"
repeat_region 37230..37345
/note="58 copies 2 mer ga 74% conserved"
repeat_region 37347..38783
/note="LIMB7 repeat: matches 2380..3857 of consensus"
repeat_region 38787..39989
/note="LIMB7 repeat: matches 3996..5195 of consensus"
repeat_region 39996..40110
/note="FLAM C repeat: matches 7..121 of consensus"
repeat_region 40111..40144
/note="BC200 repeat: matches 7..158 of consensus"
repeat_region 40145..40300
/note="LIMB7 repeat: matches 5200..5366 of consensus"
repeat_region 40304..40918
/note="LIMB7 repeat: matches 5527..6165 of consensus"
misc_feature 41788..42279
/note="match: GSS: Em:AQ587336"
repeat_region 42229..42641
/note="MIR repeat: matches 1..443 of consensus"
repeat_region 43403..43497
/note="MIR repeat: matches 84..190 of consensus"
repeat_region 43514..43711
/note="MIR repeat: matches 3..199 of consensus"
repeat_region 46007..46019
/note="MIR repeat: matches 16..28 of consensus"
repeat_region 46020..46313
/note="ALUS repeat: matches 1..296 of consensus"
repeat_region 46314..46522
/note="MIR repeat: matches 28..228 of consensus"
repeat_region 46610..46822
/note="ALU repeat: matches 14..226 of consensus"
repeat_region 46823..47188
/note="THEL1 repeat: matches 1..371 of consensus"
repeat_region 47189..47274
/note="ALU repeat: matches 226..311 of consensus"
repeat_region 48568..48726

```

```

repeat_region 49609..49812
/note="LIMC4 repeat: matches 6635..6820 of consensus"
repeat_region 49813..50168
/note="THEL1 repeat: matches 1..364 of consensus"
repeat_region 50169..50306
/note="LIMC4 repeat: matches 6820..7007 of consensus"
repeat_region 50307..50892
/note="LIMC4 repeat: matches 7168..7780 of consensus"

Query Match 2.8%; Score 73; DB 9; Length 151603;
Best Local Similarity 100.0%; Pred. No. 8.4e-30;
Matches 73; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTGACCTGAGTATCCACGACCTGCGCCCTCCCAAGTGGGATTACAG 2109
|||||
Db 69231 TCGAAGCTGACCTGAGTATCCACGACCTGCGCCCTCCCAAGTGGGATTACAG 69172
|||||

QY 2110 GTGTGAGCCACCG 2122
|||||
Db 69171 GTGTGAGCCACCG 69159
|||||

RESULT 14
AC092450_1
WPCOMMENT
Sequence split into 8 fragments
Fragment Name Begin End
AC092450_0 1 110000
AC092450_1 100001 210000
AC092450_2 200001 310000
AC092450_3 300001 410000
AC092450_4 400001 510000
AC092450_5 500001 610000
AC092450_6 600001 710000
AC092450_7 700001 727300

Continuation (2 of 8) of AC092450 from base 100001 (AC092450 Homo sapiens chromosome 12

Query Match 2.8%; Score 72; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 3.2e-29;
Matches 72; Mismatches 0; Indels 0; Gaps 0;

QY 2039 CCAGGCTGCTGCTGAACCTGACCTGAGTATCCACGACCTTGGCTCCCAAGTGC 2098
|||||
Db 97888 CCAGGCTGCTGCTGAACCTGACCTGAGTATCCACGACCTTGGCTCCCAAGTGC 97947
|||||

QY 2099 TGGGATTACAG 2110
|||||
Db 97948 TGGGATTACAG 97959
|||||

RESULT 15
AL356532/c
LOCUS
DEFINITION Human DNA sequence from clone RP3-503M14 on chromosome 1. Contains
Esrrs, GSSs and Cpg islands. Contains the 5' part of the MZPL1 gene
for myelin protein zero-like 1, complete sequence.
ACCESSION AL356532
VERSION AL356532.9 GI:10120062
KEYWORDS HTG; Cpg island; MZPL1; myelin protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 24706)
AUTHORS Coville,G.
TITLES Direct Submision
JOURNAL Submitted (19-Dec-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Sep 13, 2000 this sequence version replaced gi:9931783.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations

```



```
repeat_region      /rpt family="MIR"  
                    complement(1556..1851)  
repeat_region      /rpt family="AluBx"  
                    complement(2723..3090)
```

```

repeat_region /rpt_family="LINE2"
complement(3219..3430)
/rpt_family="LINE2"
repeat_region complement(3907..4016)
/rpt_family="(GGA)n"
repeat_region 4116..4273
/rpt_family="LINE2"
repeat_region complement(4356..4646)
/rpt_family="ALUc"
repeat_region complement(5377..5676)
/rpt_family="AluY"
repeat_region 5861..6011
/rpt_family="LINE2"
complement(6042..6126)
/rpt_family="MER5B"
repeat_region 6179..6545
/rpt_family="LINE2"
repeat_region 6779..7078
/rpt_family="AluSx"
repeat_region 7775..8156
complement(8685..8765)
/rpt_family="MTR"
repeat_region complement(10030..10749)
/rpt_family="LINE2"
repeat_region 10887..10976
/rpt_family="LINE2"
repeat_region 11106..11402
/rpt_family="AluJo"
complement(11730..11805)
/rpt_family="LINE2"
repeat_region 12330..13004
/rpt_family="LINE2"
repeat_region 13043..13188
/rpt_family="MLT1A2"
complement(13189..13490)
/rpt_family="AluSg"
repeat_region 13528..13678
/rpt_family="MLT1A2"
complement(13692..13740)
/rpt_family="L1MB7"
repeat_region 13742..14034
/rpt_family="AluSc"
complement(14035..14130)
/rpt_family="L1MB8"
complement(14130..14469)
/rpt_family="L1MB8"
repeat_region complement(14480..14781)
/rpt_family="AluJo"
repeat_region 14803..15004
/rpt_family="LINE2"
complement(15068..15155)
/rpt_family="L1PA15"
repeat_region complement(15162..15445)
/rpt_family="AluSg"
complement(15639..15724)
/rpt_family="(TA)n"
complement(15725..16021)
/rpt_family="AluSg"
repeat_region complement(16145..16535)
/rpt_family="MSTD"
repeat_region 16592..16832
/rpt_family="AluY"
repeat_region 16833..16976
/rpt_family="L1MC1"
repeat_region 17330..17641
/rpt_family="AluJo"
repeat_region 17670..17835
/rpt_family="FRAM"
complement(17924..18232)
/rpt_family="AluSx"

```

```

repeat_region 18243..18771
/rpt_family="LINE2"
complement(18325..18444)
/note="predicted exon, program: Graal2exons_human_1.3,
frame: 1, quality: good, score: 62.000"
repeat_region 18797..19097
/rpt_family="AluSx"
complement(19138..19370)
/rpt_family="MLT1C"
repeat_region 19397..19610
/rpt_family="AluSg"
complement(19618..19804)
/rpt_family="MLT1C"
complement(19863..19980)
/rpt_family="FLAM_C"
complement(20138..20477)
/rpt_family="L1MC4"
repeat_region 20478..20782
/rpt_family="AluSg"
repeat_region 20804..21109
/rpt_family="AluSx"
complement(21224..21337)
/rpt_family="AluJo"
complement(21768..22315)
/rpt_family="MER51B"
repeat_region 22391..22686
/rpt_family="AluSx"
complement(22695..22891)
/rpt_family="L1MC/D"
complement(22927..23211)
/rpt_family="AluSx"
repeat_region 23383..23497
/rpt_family="MIR"
complement(23633..23937)
/rpt_family="AluSx"
complement(24057..24291)
/rpt_family="LINE2"
repeat_region 24536..24645
/rpt_family="(TA)n"
complement(24687..24926)
/rpt_family="AluJo"
repeat_region complement(25086..25369)
/rpt_family="AluSg"
repeat_region 25466..25764
/rpt_family="AluSc"
complement(25907..26007)
/rpt_family="MIR"

```

Query Match 2 8% Score 71: DB 9: Length 38875;
 Best Local Similarity 100.0%; Pred. No. 1.3e-28; Indels 0; Gaps 0;
 Matches 71; Conservative 0; Mismatches 0;

```

QY 2050 TCGAATCTGCTGAGTGTGATCCACCCAGCTTGCCAAAGTCTGGATTACAG 2109
Db 13409 TCGAATCTGCTGAGTGTGATCCACCCAGCTTGCCAAAGTCTGGATTACAG 13468
QY 2110 GTGTGAGCCAC 2120
Db 13469 GTGTGAGCCAC 13479

```

RESULT 18
 AL596188/c 69676 bp DNA linear PRI 13-SEP-2001
 LOCUS Human DNA sequence from clone CTA-3109 on chromosome 6, complete
 DEFINITION sequence.
 ACCESSION AL596188
 VERSION AL596188.6 GI:15620698
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 69676)
 AUTHORS Tracey, A.
 TITLE Direct Submission
 JOURNAL Submitted (13-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT

On Sep 14, 2001 this sequence version replaced GI:15591680. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> CTA-31J9 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218. VECTOR: PBAC1081. IMPORTANT: This sequence is not the entire insert of clone CTA-31J9 It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-57H15 is at 2000 in this sequence.

FEATURES

source 1. 69676
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="CTA-31J9"
 /clone_id="CTG78SK-A1"
 18807..19234
 /note="CpG island"
 /evidence=not experimental
 misc_feature 60040..60136
 /note="Single clone region. Sequence from reads from a short insert library derived from a single PUC clone. Restriction digest data confirm the assembly."
 BASE COUNT 21399 a 13905 c 14237 g 20135 t
 ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 69676;
 Best Local Similarity 100.0%; Pred. No. 1.3e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGACCTGATCCACCCAGCTCCCAAGCTGGGATTACG 2109

Db 56824 TCGAAGCTCTGACCTGATCCACCCAGCTCCCAAGCTGGGATTACG 56765

QY 2110 GTGTAGCCAC 2120

Db 56764 GTGTAGCCAC 56754

RESULT 19
 AL445187/c 76114 bp DNA linear PRI 05-DEC-2001
 LOCUS Human DNA sequence from clone RP11-576C12 on chromosome 9 Contains
 DEFINITION a VP833B pseudogene, complete sequence.
 ACCESSION AL445187

VERSION AL445187.7 GI:11544549
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 76114)

COMMENT Direct Submission
 Submitted (05-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

REFERENCE

AUTHORS

TITLE

JOURNAL

On Dec 4, 2000 this sequence version replaced GI:11340318. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-576C12 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: PBAC3.6 IMPORTANT: This sequence is not the entire insert of clone RP11-576C12 it may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-576C12 is at 76114 in this sequence. The true left end of clone RP11-276E15 is at 11501 in this sequence. The true right end of clone RP11-32M23 is at 100 in this sequence.

FEATURES

SOURCE

repeat_region 1. 76114
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-576C12"
 /clone_id="RP11-576C12"
 1..232
 /note="LM4 repeat: matches 3850..4086 of consensus"
 1049..1244
 /note="MIR repeat: matches 19..256 of consensus"
 1478..1607
 /note="AluY repeat: matches 1..133 of consensus"
 2107..2406
 /note="AluY repeat: matches 1..300 of consensus"
 2696..2818
 /note="AluY repeat: matches 153..289 of consensus"
 3445..3700
 /note="MIR repeat: matches 231..488 of consensus"
 3881..4188
 /note="AluY repeat: matches 1..306 of consensus"
 4805..5110
 /note="AluY repeat: matches 1..301 of consensus"
 5126..5171
 /note="23 copies 2 mer aa 87% conserved"
 5451..5768
 /note="match: GSS: Em:AQ018476"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```

misc_feature      complement(5467..5821)
                   /note="match: GSS: Em:AQ128048"
repeat_region     5825..6127
                   /note="AluSq repeat: matches 1..301 of consensus"
misc_feature      complement(6173..6510)
                   /note="match: GSS: Em:AQ347260"
misc_feature      complement(6173..6457)
                   /note="match: GSS: Em:AQ698025"
misc_feature      6466..7152
                   /note="match: GSS: Em:AQ350137"
repeat_region     6892..7434
                   /note="MLT2D repeat: matches 9..553 of consensus"
repeat_region     7438..7556
                   /note="HERVL repeat: matches 5643..5761 of consensus"
repeat_region     7665..7834
                   /note="MLT1J repeat: matches 1..180 of consensus"
repeat_region     7877..7999
                   /note="FLAM_C repeat: matches 5..130 of consensus"
repeat_region     8001..8086
                   /note="43 copies 2 mer aa 76% conserved"
repeat_region     8943..9042
                   /note="L2 repeat: matches 2214..2313 of consensus"
misc_feature      9154..9172
                   /note="Single clone region. Assembly confirmed by
restriction digest data."
misc_feature      9173..9245
                   /note="Single clone region. Assembly confirmed by
restriction digest data."
misc_feature      complement(11049..11541)
                   /note="match: GSS: Em:AQ678143"
repeat_region     12551..13531
                   /note="LIMB7 repeat: matches 5155..6173 of consensus"
misc_feature      complement(13184..13755)
                   /note="match: GSS: Em:B47740"
misc_feature      13772..14259
                   /note="match: GSS: Em:AQ608809"
repeat_region     15365..15813
                   /note="LIMB7 repeat: matches 4783..5233 of consensus"
repeat_region     16572..16647
                   /note="38 copies 2 mer at 72% conserved"
repeat_region     16711..17016
                   /note="AluSq repeat: matches 3..308 of consensus"
repeat_region     17295..17505
                   /note="MIR repeat: matches 34..250 of consensus"
repeat_region     17626..17696
                   /note="MER5A repeat: matches 12..79 of consensus"
gene              complement(join(17983..18085,18189..18257,18683..18794))
                   /gene="bA576C12.1"
CDS               complement(join(17983..18085,18189..18257,18683..18794))
                   /gene="bA576C12.1"
                   /note="bA576C12.1 (VPS33B pseudogene)
match: proteins: Tr:Q9NSF3 Tr:Q63616"
pseudo           /codon_start=1
evidence=not_experimental
repeat_region     18376..18594
                   /note="LIM2 repeat: matches 6088..6308 of consensus"
repeat_region     19078..19134
                   /note="L2 repeat: matches 2647..2706 of consensus"
repeat_region     19152..19244
                   /note="MER5A repeat: matches 87..189 of consensus"
repeat_region     19623..19833
                   /note="LIMC/D repeat: matches 5412..5801 of consensus"
repeat_region     20152..20280
                   /note="AluJo/PRAM repeat: matches 165..293 of consensus"
repeat_region     20325..20542
                   /note="MIR repeat: matches 7..238 of consensus"
repeat_region     22120..22248
                   /note="L2 repeat: matches 2619..2750 of consensus"
repeat_region     23364..23653
                   /note="AluX repeat: matches 1..292 of consensus"
repeat_region     24037..24135

```

```

repeat_region     24538..24614
                   /note="HAI repeat: matches 1434..1535 of consensus"
repeat_region     25175..25283
                   /note="MIR repeat: matches 168..247 of consensus"
repeat_region     25519..25798
                   /note="FLAM_C repeat: matches 1..109 of consensus"
repeat_region     26291..26428
                   /note="AluSD repeat: matches 1..295 of consensus"
repeat_region     26552..26854
                   /note="MER5A repeat: matches 15..184 of consensus"
repeat_region     27292..27587
                   /note="AluSq repeat: matches 1..304 of consensus"
repeat_region     27815..28006
                   /note="Alu repeat: matches 1..295 of consensus"
repeat_region     28021..28470
                   /note="Alu repeat: matches 1..192 of consensus"
repeat_region     28478..28582
                   /note="LTP10A repeat: matches 13..548 of consensus"
repeat_region     28585..28812
                   /note="Alu repeat: matches 188..292 of consensus"
repeat_region     29785..29914
                   /note="MIR repeat: matches 8..256 of consensus"
repeat_region     30054..30368
                   /note="FLAM_C repeat: matches 1..132 of consensus"
repeat_region     30465..31203
                   /note="AluJo repeat: matches 1..312 of consensus"
repeat_region     31583..31900
                   /note="LIMB7 repeat: matches 5377..6168 of consensus"
repeat_region     31963..31963
                   /note="AluX repeat: matches 1..301 of consensus"
repeat_region     31995..32288
                   /note="Sequence from overlapping clone bA276B15
(AU390067). Assembly confirmed by restriction digest."
repeat_region     33495..33681
                   /note="MIR repeat: matches 1..295 of consensus"
repeat_region     33999..34064
                   /note="match: GSS: Em:AQ743239"
repeat_region     34080..34720
                   /note="33 copies 2 mer tc 66% conserved"
repeat_region     34993..35270
                   /note="L2 repeat: matches 2080..2710 of consensus"
repeat_region     37056..37286
                   /note="MIR repeat: matches 1..274 of consensus"
repeat_region     37287..37427
                   /note="match: GSS: Em:AQ481167"
repeat_region     37287..37427
                   /note="MIR repeat: matches 1..232 of consensus"
repeat_region     37287..37427
                   /note="L2 repeat: matches 1267..1415 of consensus"

```

```

Query Match      2.8%; Score 71; DB 9; Length 76114;
Best Local Similarity 100.0%; Prid. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2050 TCGAATCTTACCTGAGTGATCCACCCACCTTGCTCCCAAGTGTGGATTACAG 2109
DB 31664 TCGAATCTTACCTGAGTGATCCACCCACCTTGCTCCCAAGTGTGGATTACAG 31605

Oy 2110 GTGTGAGCCAC 2120
DB 31604 GTGTGAGCCAC 31594

RESULT 20
AL354986/c 88557 bp DNA linear HTG 10-Jul-2001
LOCUS AL354986 Homo sapiens chromosome 1 clone R95-1100D14 map p34.1-35.***
DEFINITION SEQUENCING IN PROGRESS ***, 8 unordered pieces.
ACCESSION AL354986
VERSION AL354986.4 GI:11229132
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.

```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 88557)
 AUTHORS McLay, K.
 TITLE Direct Submission
 JOURNAL Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Nov 20, 2000 this sequence version replaced gi:9797233.

COMMENT

----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

----- Project Information
 Center project name: d1100D14

----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator ABI; 2% of reads
 Chemistry: Dye-terminator Big Dye; 84% of reads
 Chemistry: Dye-terminator ET-amersham; 13% of reads
 Quality: 8653 bases at least Q40
 Consensus quality: 87145 bases at least Q30
 Consensus quality: 87857; sum-of-contigs
 Insert size: 109672; 31.9% error; agarose-fp
 Insert size: 109672; 31.9% error; agarose-fp
 Quality coverage: 5.81x in Q20 bases; sum-of-contigs quality
 coverage: 4.81x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 8490: contig of 8490 bp in length
 * 8491 8590: gap of 100 bp
 * 8591 40289: contig of 31709 bp in length
 * 40300 40399: gap of 100 bp
 * 40400 45544: contig of 5145 bp in length
 * 45545 45644: gap of 100 bp
 * 45645 49137: contig of 3493 bp in length
 * 49138 49237: gap of 100 bp
 * 49238 53403: contig of 4166 bp in length
 * 53404 53503: gap of 100 bp
 * 53504 78346: contig of 24643 bp in length
 * 78347 78446: gap of 100 bp
 * 78447 86288: contig of 7842 bp in length
 * 86289 86388: gap of 100 bp
 * 86389 88557: contig of 2169 bp in length.
 Location/Qualifiers
 1..88557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p34.1-35"
 /clone="RP5-1100D14"
 /clone_id="RPC1-5"
 1..8490
 /note="assembly_fragment:00310
 fragment_chain:1"
 8591..40289
 /note="assembly_fragment:00014
 fragment_chain:1"
 40400..45544
 /note="assembly_fragment:00898
 fragment_chain:2"
 45645..49137
 /note="assembly_fragment:00104

FEATURES

source

misc_feature

/note="assembly_fragment:00310
 fragment_chain:1"

misc_feature

/note="assembly_fragment:00014
 fragment_chain:1"

misc_feature

/note="assembly_fragment:00898
 fragment_chain:2"

misc_feature

/note="assembly_fragment:00104

misc_feature
 49238..53403
 /note="assembly_fragment:00712"
 53504..78346
 /note="assembly_fragment:01219"
 78447..86288
 /note="assembly_fragment:00674
 fragment_chain:3"
 86389..88557
 /note="assembly_fragment:01199
 fragment_chain:3
 clone end:17
 vector side:right"

BASE COUNT 22574 a 21575 c 20359 g 23347 t 702 others
 ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 88557;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTCGACCTGAGTATCCACCCCTGAGCTCCCAAGTGTGGATTACAG 2109
 Db 68855 TCGAAGCTCGACCTGAGTATCCACCCCTGAGCTCCCAAGTGTGGATTACAG 68796

Qy 2110 GTGTGACCCAC 2120
 Db 68795 GTGTGACCCAC 68785

RESULT 21

AC090114 91722 bp DNA linear PRI 13-MAR-2002
 LOCUS Homo sapiens BAC clone RP11-212P7 from 7, complete sequence.
 DEFINITION AC090114
 ACCESSION AC090114.5 GI:19387791
 VERSION HTG.
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 91722)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE 2 (bases 1 to 91722)
 AUTHORS Maligorski, J., Meyer, R. and Doeber, A.
 TITLE The sequence of Homo sapiens BAC clone RP11-212P7
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 91722)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (16-FEB-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 4 (bases 1 to 91722)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Mar 13, 2002 this sequence version replaced gi:18151035.

COMMENT

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0212P07

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatenio, M., Calanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-155614, 2000 bp overlap. The clone sequenced to the right is RP13-485A5. Actual end of this clone is at base position 91722 of RP11-212P7.

FEATURES

source
1. 91722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-212P7"
/clone_id="RPCI-11"
1. 166
/rpt_family="Alu"
167..216
/rpt_family="Mariner"
235..347
/rpt_family="Alu"
453..606
/note="match to EST AL553530 (NID:g12893451)"
453..602
/note="match to EST AW957421 (NID:GB147224)"
453..602
/note="match to EST BF033256 (NID:g10740968)"
453..602
/note="similar to Mus musculus EST BG069684 (NID:g12552253)"
457..602
/note="similar to Homo sapiens EST AL520867 (NID:g12784360)"
460..602
/note="similar to EST BM452818 (NID:g18501858)"
475..602
/note="match to EST BG181401 (NID:g13703088)"
502..602
/note="match to EST BF332291 (NID:g11303039)"
526..602
/note="match to EST BF332291 (NID:g11303039)"
530..602
/note="similar to Bos taurus EST BM107596 (NID:g17038666)"
550..588
/note="match to EST AU118175 (NID:g10933201)"
misc_feature
/note="similar to Homo sapiens EST AL638878 (NID:g4691112)

repeat_region
600..950
/rpt_family="Alu"
repeat_region
1286..1577
/rpt_family="Alu"
repeat_region
1583..1879
/rpt_family="Alu"
1803..2089
/note="match to EST BF032110 (NID:g12180498)"
1949..1952
/note="match to EST N33555 (NID:g1153954) YY04f07.s1"
2102
/note="match to EST BF032110 (NID:g12180498)"
2103..2411
/rpt_family="Alu"
2341..2699
/note="similar to Homo sapiens EST AW812325 (NID:g7805319)"
2435..2611
/note="similar to Homo sapiens EST BF328761 (NID:g1299509)"
2588..2592
/note="match to EST AW352175 (NID:g6849888)"
2646..2647
/note="similar to Homo sapiens EST BF328761 (NID:g1299509)"
2922..3215
/rpt_family="Alu"
3275..3295
/rpt_family="AT-rich"
3299..3394
/rpt_family="Alu"
3416..3454
/rpt_family="(TTTG)n"
3457..3697
/rpt_family="Alu"
3725..4207
/note="match to EST N33555 (NID:g1153954) YY04f07.s1"
3943..3946
/note="similar to Homo sapiens EST BG403117 (NID:g13296565)"
3956..4016
/rpt_family="L2"
4134..4441
/rpt_family="Alu"
4443..4763
/note="match to EST BE068262 (NID:g8412912)"
4678..4801
/note="similar to Homo sapiens EST AL638878 (NID:g4691112) t108d05.x1"
4680
/note="match to EST H98925 (NID:g1123593) YX31a12.s1"
4684..4801
/note="match to EST BE792117 (NID:g10213315)"
4693..5154
/note="match to EST H98925 (NID:g1123593) YX31a12.s1"
4695..4840
/note="match to EST AW957421 (NID:g8412924)"
4695..4803
/note="match to EST BF332291 (NID:g11303039)"
4695..4801
/note="match to EST AU118175 (NID:g10933201)"
4695..4801
/note="match to EST BF033256 (NID:g10740968)"
4695..4801
/note="match to EST BG181401 (NID:g13703088)"
4695..4801
/note="similar to Bos taurus EST BM107596 (NID:g17038666)"
4695..4801
/note="similar to EST BM452818 (NID:g18501858)"
4695..4801
/note="similar to Homo sapiens EST AL520867 (NID:g12784360)"

TITLE Direct Submission
JOURNAL Submitted (15-APR-1999) Mika Hirakawa, Japan Science and Technology Corporation (JST), Advanced Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081, Japan (E-mail: mika@okyo.jst.go.jp, URL: http://www-alis.tokyo.jst.go.jp/, Tel: 81-3-5214-8491, Fax: 81-3-5214-8470)

COMMENT This sequence is submitted by Human Genome Sequencing in ALIS project of JST.
 Laboratory as a JST sequencing team.
 Principal Investigator: Yoshiyuki Sakaki Ph.D.
 Phone: +81-3-5449-5622, Fax: +81-3-5449-5445, sakaki@qcc.ims.u-tokyo.ac.jp
 Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The sequence is submitted by Human Genome Sequencing in ALIS project of JST.

FEATURES
 source
 1..100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="21q22.1"
 /chromosome="21"
 14148..14271
 /note="D55705; The location is between each flanking site of PCR primers."
 /db_xref="GDB:6458824"
 14257..14410
 /note="SHGC-51915; The location is between each flanking site of PCR primers."
 /db_xref="GDB:6464637"
 28193..28286
 /standard_name="D21S63"
 /note="KM1712/KM2830; The location is between each flanking site of PCR primers."
 /db_xref="GDB:287574"

BASE COUNT 24910 a 22647 c 23866 g 28977 t

ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 100000;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAATCTCGACCTGAGTATCCACCCAGCTGCTCCCAAGTCTGGGATTACAG 2109
 TCGAATCTCGACCTGAGTATCCACCCAGCTGCTCCCAAGTCTGGGATTACAG 34700

DB 34759

QY 2110 GTGTGAGCCAC 2120
 GTGTGAGCCAC 34689

RESULT 24
AP000174/c
LOCUS AP000174 100000 bp DNA linear PRI 08-JAN-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
 clone Q78C10-f32e9, segment 1/21, complete sequence.
ACCESSION AP000174
VERSION AP000174.1 GI:4827139
KEYWORDS HTG.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 100000)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., and Sakaki, Y.
TITLE Homo sapiens 2,083,744bp genomic DNA of 21q22.1 (REGION:
 D21S226-AML CLONE RANGE: Q78C10-f32e9)
JOURNAL Published Only in Database (1999)

REFERENCE 2 (bases 1 to 100000)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail: hattori@gsc.riken.go.jp,
 URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
 Fax: 81-42-778-9924)

COMMENT E. coli transposon insertion: The present data does not contain E.
 coli transposon sequences which integrated in the
 original/previous sequences. We determined the boundary between
 the insertion and genomic sequences experimentally, removed the
 insertion sequences, reconstructed the present data. The sequencing
 project is supported by Japan Science Technology Corporation (JST)
 and The Institute of Physical and Chemical Research (RIKEN).

FEATURES
 source
 1..100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="21q22.1"
 /chromosome="21"

BASE COUNT 24910 a 22647 c 23866 g 28977 t

ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 100000;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAATCTCGACCTGAGTATCCACCCAGCTGCTCCCAAGTCTGGGATTACAG 2109
 TCGAATCTCGACCTGAGTATCCACCCAGCTGCTCCCAAGTCTGGGATTACAG 34700

DB 34759

QY 2110 GTGTGAGCCAC 2120
 GTGTGAGCCAC 34689

RESULT 25
AL365505
LOCUS AL365505 101500 bp DNA linear PRI 14-MAR-2001
DEFINITION Human DNA sequence from clone RP11-382A12 on chromosome 20 Contains
 ESTs, STSs, GSSs and CpG islands. Contains the 5' part of the gene
 for a 72.1 kDa protein (DKFZP564A032, SBBI88) similar to mouse
 IFN-gamma induced MG11 and the 3' part of the RB11 gene for
 retinoblastoma-like protein 1 (p107), complete sequence.
ACCESSION AL365505
VERSION AL365505.15 GI:11568572
KEYWORDS HTG; CpG island; MG11; RB11; retinoblastoma-like.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 101500)
AUTHORS Matthews, L.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Dec 5, 2000 this sequence version replaced gi:11545056.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP, Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human

* chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/cep/Chr20>
Rpl1-382A12 is from the library RPl-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-382A12. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone Rpl1-332A4 is at 101401 in this sequence. The true right end of clone Rpl-132P21 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES

source

1. 101500

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/clone="Rpl1-382A12"

/clone_lib="RPl-11.2"

Location/Qualifiers

1. 101500

/note="AluY repeat: matches 1. .298 of consensus"

336. .520

/note="AluY repeat: matches 149. .304 of consensus"

544. .841

/note="AluX repeat: matches 1. .298 of consensus"

844. .1029

/note="MIR repeat: matches 49. .257 of consensus"

1069. .1185

/note="FLAM C repeat: matches 1. .118 of consensus"

/gene="dJ132P21.3"

complement(join(<1314. .1429,2557. .2648,6818. .6926,

7045. .7145,9460. .9615,17278. .17348,20856. .20971,

25125. .25265,31135. .31207,36834. .36900,41532. .41869))

/gene="dJ132P21.3"

/product="bA382A12.1 (72.1 kDa protein (DKFZP564A032,

SBB188) similar to mouse IFN-gamma induced MG11)"

/note="match: CDNA: Em:AF228421 Em:U15635 Em:AL050267

Em:AK024353 Em:AF147427

match: ESTs: Em:BE888325 Em:AW958545 Em:AA381682

Em:AV734982 Em:AT744621 Em:AW504769 Em:R25510 Em:AA496236

Em:AA381555 Em:MB6047 Em:AA542149 Em:BE818948 Em:BE006171

Em:BE818922 Em:W20038 Em:AA31478 Em:AA381497 Em:AA381593

Em:BE006144 Em:AA478751 Em:BE843840 Em:R1341137

Em:BE671955 Em:H47862 Em:AI675309 Em:AW277012"

/evidence=not_experimental

complement(join(<1314. .1429,2557. .2648,6818. .6926,

7045. .7145,9460. .9615,17278. .17348,20856. .20971,

25125. .25265,31135. .31207,36834. .36900,41532. .41739))

CDS

/gene="dJ132P21.3"

/note="continues in dJ132P21 (AL079335)

match: CDNA: Em:AF228421 Em:U15635 Em:AL050267

Em:AK024353 Em:AF147427

match: ESTs: Em:BE888325 Em:AW958545 Em:AA381682

Em:AV734982 Em:AT744621 Em:AW504769 Em:R25510 Em:AA496236

Em:AA381555 Em:MB6047 Em:AA542149 Em:BE818948 Em:BE006171

Em:BE818922 Em:W20038 Em:AA31478 Em:AA381497 Em:AA381593

Em:BE006144 Em:AA478751 Em:BE843840 Em:R1341137

Em:BE671955 Em:H47862 Em:AI675309 Em:AW277012

match: proteins: Tr:O91323 SW:Q6U710 SW:O09374 Tr:O9VIV4

Tr:O26251 Tr:O28849 Tr:P95966 Tr:Q58554 Tr:O9VIT1

Tr:O9XIN3 Tr:O59446 Tr:O67745"

/codon_start=1

/evidence=not_experimental

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/product="bA382A12.1 (72.1 kDa protein (DKFZP564A032,

SBB188) similar to mouse IFN-gamma induced MG11)"

/protein_id="CAC19806.1"

/db_xref="GI:12043445"

/translation="MGRADSDSPSKRPRCDSPPTSPTEAEDNSPGLEHPDYKT
WGPPQVSPFRRGFEPEVLLKNTRENETGALPCDESEFENLGVSSIGERKULS
YIOLVQIHVDTKVYNDPIGHITELPLVRIIDTPOFORLRYIKOLGGGVYVPGA
SHNFEHSLGVGLAGCLVHALGKPELDISRPVLCVQIAGLCHDGHGSPSHWFD
GRFTPLARPKVKTQSGVMMFPHLINSNGIKPVMQYGLIPEDICTFKEQIVGL
ESPVEDLMPVKGRPEKNSFLYEIVSNKRAGIDVDKMPYFARDCHGIGIONPDYKRF
IKFARVUNNEIRICARDKEVCNLYMPPHTNSLHRRAYQKXKNIDIMTDAFLK
ADDTIELTGSGKKRISTIAIDMEATKLT"

1604. .1717

/note="MIR repeat: matches 116. .238 of consensus"

2849. .2934

/note="FLAM A repeat: matches 43. .142 of consensus"

2992. .3049

/note="Alu repeat: matches 243. .300 of consensus"

3083. .3392

/note="Alu repeat: matches 1. .309 of consensus"

complement(3140. .3337)

/gene="dJ132P21.3"

/note="match: GSS: Em:A0601622"

complement(3157. .3379)

/gene="dJ132P21.3"

/note="match: GSS: Em:A0601622"

3118. .3716

/note="AluX repeat: matches 1. .300 of consensus"

3545. .3684

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

3719. .3831

/note="MIR repeat: matches 148. .254 of consensus"

3853. .4149

/note="AluX repeat: matches 1. .297 of consensus"

4241. .4441

/note="Tigzer2a repeat: matches 1. .219 of consensus"

4442. .4755

/note="AluX repeat: matches 1. .305 of consensus"

4756. .4824

/note="Tigzer2a repeat: matches 219. .293 of consensus"

4925. .5125

/note="AluY repeat: matches 5. .300 of consensus"

5126. .5157

/note="Tigzer2a repeat: matches 293. .326 of consensus"

5158. .5449

/note="AluY repeat: matches 1. .298 of consensus"

5450. .5550

/note="Tigzer2a repeat: matches 326. .435 of consensus"

5551. .5857

/note="AluX repeat: matches 1. .307 of consensus"

5858. .5927

/note="Tigzer2a repeat: matches 435. .434 of consensus"

5974. .6267

/note="AluY repeat: matches 1. .292 of consensus"

6274. .6445

/note="MIR repeat: matches 29. .230 of consensus"

6447. .6573

/note="FLAM A repeat: matches 1. .127 of consensus"

7468. .7763

/note="AluX repeat: matches 1. .297 of consensus"

8116. .8346

/note="AluY repeat: matches 1. .294 of consensus"

8433. .8560

/note="FLAM C repeat: matches 1. .128 of consensus"

8613. .8927

/note="AluX repeat: matches 1. .312 of consensus"

8951. .9240

/note="AluY repeat: matches 2. .291 of consensus"

10228. .10491


```

repeat_region      /rpt_family="MIR"
                    complement(17712..17885)
repeat_region      /rpt_family="MERSA"
                    complement(18533..18714)
repeat_region      /rpt_family="MERS20"
                    complement(21055..21124)
repeat_region      /rpt_family="L2"
                    complement(21201..21250)
repeat_region      /rpt_family="L2"
                    complement(21657..21707)
repeat_region      /rpt_family="MIR"
                    complement(22379..22649)
repeat_region      /rpt_family="LMC1"
                    complement(22979..23292)
repeat_region      /rpt_family="AluSg"
                    complement(24216..24527)
repeat_region      /rpt_family="AluO"
                    complement(24687..24980)
repeat_region      /rpt_family="L1PA4"
                    complement(25755..25782)
repeat_region      /rpt_family="AT_rich"
                    complement(25785..26070)
repeat_region      /rpt_family="AluSx"
                    complement(26071..26107)
repeat_region      /rpt_family="AT_rich"
                    complement(26186..26477)
repeat_region      /rpt_family="AluO"
                    complement(26594..26757)
repeat_region      /rpt_family="MIR"
                    complement(26902..27202)
repeat_region      /rpt_family="AluSg"
                    complement(28990..29296)
repeat_region      /rpt_family="AluO"
                    complement(29742..29896)
repeat_region      /rpt_family="MIR"
                    complement(30043..30592)
repeat_region      /rpt_family="MLT1F"
                    complement(31090..31367)
repeat_region      /rpt_family="AluSx"
                    complement(32074..32199)
repeat_region      /rpt_family="MIR"
                    complement(32206..32521)
repeat_region      /rpt_family="AluSg"
                    complement(32572..32757)
repeat_region      /rpt_family="MIR"
                    complement(32761..32838)
repeat_region      /rpt_family="MERS2"
                    complement(33407..33553)
repeat_region      /rpt_family="MERSB"
                    complement(33573..33756)
repeat_region      /rpt_family="MERSA"
                    complement(33771..33973)
repeat_region      /rpt_family="MERS8A"
                    complement(34187..34271)
repeat_region      /rpt_family="MIR"
                    complement(34985..35288)
repeat_region      /rpt_family="AluSx"
                    complement(35412..35477)
repeat_region      /rpt_family="MERSA"
                    complement(35620..36292)
repeat_region      /rpt_family="L1PA2"
                    complement(36302..36575)
repeat_region      /rpt_family="L1MB5"
                    complement(36577..36686)
repeat_region      /rpt_family="MIR"
                    complement(36933..37007)
repeat_region      /rpt_family="TGAA)n"
                    complement(37875..37999)
repeat_region      /rpt_family="FLAM_A"
                    complement(38476..38556)
repeat_region      /rpt_family="PTR5"
                    complement(38557..38849)
repeat_region      /rpt_family="AluY"

```

```

repeat_region      38850..39247
                    /rpt_family="LTR12"
repeat_region      complement(39625..39748)
                    /rpt_family="MIR"
repeat_region      40650..40674
                    /rpt_family="CAG)n"
repeat_region      42135..42188
                    /rpt_family="(GGA)n"
repeat_region      complement(43811..43983)
                    /rpt_family="MIR"
repeat_region      44313..44493
                    /rpt_family="MERSB"
repeat_region      complement(44546..44620)
                    /rpt_family="MERS8A"
repeat_region      complement(45397..45480)
                    /rpt_family="L1MB8"
repeat_region      45517..45584
                    /rpt_family="MLT1F"
repeat_region      complement(45696..45804)

```

Query Match 2.8%; Score 71; DB 9; Length 104147;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAACCTCTGACCTGAGTGATCACCACCTG3CCTCCCAAGCTGGGATTACAG 2109
 Db 90800 TCGAATCTCGACCTGAGTGATCACCACCTG3CCTCCCAAGCTGGGATTACAG 90741

Qy 2110 GGTGAGGCAC 2120
 Db 90740 GGTGAGGCAC 90730

RESULT 27
 LOCUS AC091614 107764 bp DNA linear PRI 22-DEC-2001
 DEFINITION Homo sapiens chromosome 1 clone RP4-65J23, complete sequence.
 AC091614 AL356854
 VERSION AC091614.3 GI:1977203
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 107764)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
 Haugen, E.D.
 Direct Submisson
 Submitted (09-MAY-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 107764)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
 Haugen, E.D.
 Direct Submisson
 Submitted (22-DEC-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Dec 22, 2001 this sequence version replaced gi:15145523.

REFERENCE AUTHORS
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
 Haugen, E.D.
 Direct Submisson
 Submitted (09-MAY-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 107764)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
 Haugen, E.D.
 Direct Submisson
 Submitted (22-DEC-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Dec 22, 2001 this sequence version replaced gi:15145523.

Center: University of Washington Genome Center
 Center Code: UMG
 Web site: <http://www.genome.washington.edu>
 Contact: uwgchgs@u.washington.edu
 Drafting Center: SC
 Project Information
 Center project name: chr-1
 Center clone name: RP4-65J23 (sc0203)
 Summary Statistics
 Sequencing vector: plasmid, p08752, 100% of reads

Chemistry: Dye-terminator Big Dye, 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 107697 bases at least Q40
 Consensus quality: 107756 bases at least Q30
 Consensus quality: 107764 bases at least Q20
 Insert size: 107581; sum-of-contrigs
 Quality coverage: 8.7x in Q20 bases; sum-of-contrigs

Overlapping Sequences:
 5': Mapping in progress
 3': RPS-924G13 AL136109

Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

SeqDerMap	FingerPrint	SeqDerMap	FingerPrint	SeqDerMap	FingerPrint
-----	-----	-----	-----	-----	-----
6141	6207	2186	2167	9719	9643
-----	-----	-----	-----	-----	-----
2184	2186	449	<800	5671	5706
-----	-----	-----	-----	-----	-----
8065	8225	512	<800	4252	4269
-----	-----	-----	-----	-----	-----
486	<800	2814	2829	4573	4548
-----	-----	-----	-----	-----	-----
106	<800	1247	1203	1388	1389
-----	-----	-----	-----	-----	-----
5839	5881	4695	4744	8582	8585
-----	-----	-----	-----	-----	-----
5024	5124	3373	3404	2326	2507
-----	-----	-----	-----	-----	-----
843	843	3940	3959	12131	12076
-----	-----	-----	-----	-----	-----
15970	15803	976	964	1392	1389
-----	-----	-----	-----	-----	-----
2767	2784	513	<800	1410	1389
-----	-----	-----	-----	-----	-----
676	<800	612	<800	7569	7617
-----	-----	-----	-----	-----	-----
19591	19553	5767	5877	6505	6523
-----	-----	-----	-----	-----	-----
23557	23906	2059	2044	2944	2954
-----	-----	-----	-----	-----	-----

-----	1568	1551	850	862	6283	6264
-----	5999	5881	10247	10174	580	<800
-----	5142	5124	4011	3959	743	748
-----	121	<800	3566	3557	23	<800
-----	18437	18067	872	862	14146	14238
-----	1250	1234	4776	4744	2394	2507
-----	5883	5877	1672	1672	1616	
-----	10636	10530	3196	3196	3125	
-----	8601	8587	236	<800	<800	
-----	2088	2167	2507	2507	2507	
-----	2110	21186	1312	1295		
-----	7406	7332	5546	5472		
-----	241	<800	2508	2507		
-----	3917	3959	1048	1034		
-----	859	862	4851	4814		
-----	166	<800	4330	4269		
-----	3133	3108	3594	3583		
-----	2682	2707	335	<800		
-----	392	<800				
-----	2814	2829				
-----	373	<800				

FEATURES
 source location/Qualifiers
 1..107764
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP4-665J23"
 /clone_1fb="RPC1 human PAC library 4"
 45711..45818
 /note="single subclone region"
 BASE COUNT 3346 a 21787 c 20575 g 32056 t
 ORIGIN
 Query Match 2.8%; Score 71; DB 9; Length 107764;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2050 TCGAATCTGACCTGAGTATCCACCCAGCTGCGCTCCCAAGTGGATTACAG 2109
 DB 49247 TCGAATCTGACCTGAGTATCCACCCAGCTGCGCTCCCAAGTGGATTACAG 49306
 QY 2110 GTGTGAGCCAC 2120
 DB 49307 GTGTGAGCCAC 49317
 RESULT 28
 AC131236/c 108239 bp DNA linear HTG 19-AUG-2002
 LOCUS AC131236 Homo sapiens clone CTD-2260B7, *** SEQUENCING IN PROGRESS ***, 14
 DEFINITION unorderd pieces.
 ACCESSION AC131236

VERSION
AC131236.1 GI:22297331
KEYWORDS
HTG, HTGS_PHASE1.
SOURCE
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Muzny D.M., Adams C.C., Adio-Oduola B., Alt-omani F.R., Allen C., Alibekova S.L., Amaralunga H.C., Are J.R., Ayala M., Banks T., Barbara J., Benton J., Bimaga K., Blankenship K., Bonin D., Bouck J., Bowe S., Brieva M., Brown E., Brown M., Bryant N.P., Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C., Caron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C., Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O., Derra A.L., Ding X., Dinh H.H., Doukwaile K.J., Draper H., Dugan-Rocha S., Durbin K.U., Eamhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Franz P., Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K., Harrie C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B., Hones F., Howard S., Huber J., Huijck S., Hume J., Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S., Joudan S., Karlsson E., Kelly S., Khan U., King L., Koryak J., Kovar C., Kratovic J., Kuresh A., Landry N., Leal B., Lewis L.C., Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loussed H., Lozada R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J., Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E., Maeser E., Maehney E., McLeod T., Mohabbat K., Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokkwo S., Ogih M., Okwunodu G., Ogunye N., Oyedero R., Pace A., Payton B., Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y., Rivers M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savary G., Scherer S., Scott G., Shen H., Shooshitari N., Sisson I., Sodergren E., Sonaike T., Sparks A., Stanley H., Stone H., Sutton A., Svatek A., Tabor P., Tamerisa K., Tamerisa K., Tang H., Tansey J., Taylor C., Taylor T., Telirod B., Thomas N., Thomas S., Usman K., Vaequez U., Vera V., Villalon D., Vinson R., Wang Q., Wang S., Ward-Moore S., Warren R., Washington C., Watlington S., Williams G., Williamson A., Wleczek R., Woodson S., Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D., Weinstein G., and Gibbs R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT
Unpublished
2 (bases 1 to 108239)
Worley K.C.
Direct Submission
Submitted (19-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HELI
Center clone name: CTD-2260B7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 1000 of reads
Assembly program: Phrap, version 0.990129
Consensus quality: 97894 bases at least Q40
Consensus quality: 101040 bases at least Q30
Consensus quality: 103062 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2543: contig of 2543 bp in length
2544 2643: gap of unknown length
2544 5633: contig of 2590 bp in length
5634 5733: gap of unknown length
5734 9343: contig of 3610 bp in length
9344 9444: gap of unknown length
9444 12821: contig of 3378 bp in length
12822 12921: gap of unknown length
12922 18419: contig of 5498 bp in length
18420 18519: gap of unknown length
18520 25079: contig of 6560 bp in length
25080 25179: gap of unknown length
25180 35297: contig of 10118 bp in length
35298 35397: gap of unknown length
35398 42548: contig of 7151 bp in length
42549 42648: gap of unknown length
42649 52359: contig of 9611 bp in length
52360 52359: gap of unknown length
52360 65801: contig of 13442 bp in length
65802 72723: gap of unknown length
72723 72724: contig of 6822 bp in length
72724 72823: gap of unknown length
72824 82701: contig of 9884 bp in length
82701 82708: gap of unknown length
82708 95092: contig of 12285 bp in length
95092 95192: gap of unknown length
95192 108239: contig of 13047 bp in length.
Location/Qualifiers
1..108239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CMD-2260B7"

BASE COUNT 28670 a 26415 c 24912 g 26698 t 1544 others
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 108239;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTGACCTGACGATGATCCACCCCTGCTCCCAAGCTGGATTACAG 2109
|||||
Db 3860 TCGAAGCTGACCTGACGATGATCCACCCCTGCTCCCAAGCTGGATTACAG 3801
QY 2110 GTGTGAGCCAC 2120
|||||
Db 3800 GTGTGAGCCAC 3790

RESULT 29
AL390208/c
LOCUS 109878 bp DNA linear PRI 20-JAN-2001
DEFINITION Human DNA sequence from clone Rp11-787122 on chromosome 6, complete
sequence.
ACCESSION AL390208
VERSION AL390208.17 GI:12331104
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 109878)
AUTHORS Blakey, S.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

COMMENT

request: clonerequest@sanger.ac.uk
 On Jan 22, 2001 this sequence version replaced g1:12329420.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated repeat sequence elements. Where the sequence is
 ambiguous, there is an annotation using the 'unsure' feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP11-787122 is from the library RP11-11.3 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-787122. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-787122 is at 1 in this sequence.
 The true left end of clone RP11-48723 is at 109779 in this
 sequence. The true right end of clone RP3-397N18 is at 97339 in
 this sequence.

FEATURES

source
 1. 109878
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-787122"
 /clone_1fb="RP11-11.3"
 /clone_1fb="RP11-11.3"
 misc_feature
 8165_81200
 /note="Sequence from AC009282. Sequenced by Whitehead."
 BASE COUNT 31786 a 21771 c 21950 g 34371 t
 ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 109878;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2050 TCGAATCTCTAGCTAGGTGATCCACCCACTTGGGCTCCCAAGTCTGGGATTACAG 2109

DB 44625 TCGAATCTCTAGCTAGGTGATCCACCCACTTGGGCTCCCAAGTCTGGGATTACAG 44566

OY 2110 GTGTGAGCCAC 2120

DB 44565 GTGTGAGCCAC 44555

RESULT 30
 AC107310 2/c
 WPCOMMENT

Sequence split into 4 fragments LOCUS AC107310 Accession AC107310

Fragment Name Begin End
 AC107310_0 1 110000
 AC107310_1 100001 210000
 AC107310_2 200001 310000
 AC107310_3 300001 357311

Continuation (3 of 4) of AC107310 from base 200001 (AC107310 Homo sapiens chromosome 3 c

Query Match 2.8%; Score 71; DB 2; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2050 TCGAATCTCTAGCTAGGTGATCCACCCACTTGGGCTCCCAAGTCTGGGATTACAG 2109
 DB 44603 TCGAATCTCTAGCTAGGTGATCCACCCACTTGGGCTCCCAAGTCTGGGATTACAG 44544
 OY 2110 GTGTGAGCCAC 2120
 DB 44543 GTGTGAGCCAC 44533

RESULT 31

AL357500
 LOCUS AL357500 112392 bp DNA linear PRI 15-NOV-2001
 DEFINITION Human DNA sequence from clone RP11-242024 on chromosome 1, complete
 sequence.
 ACCESSION AL357500
 VERSION AL357500.17 GI:16972897
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 112392)
 Bates, K.
 Direct Submission
 Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humanyes@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Nov 16, 2001 this sequence version replaced g1:16944735.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

COMMENT

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
 RP11-242024 is from the library RP11-11.1 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-242024. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true right end of clone RP11-242024 is at 112392 in this
 sequence. The true right end of clone RP4-604A21 is at 2000 in this
 sequence.

FEATURES

source
 1. 112392
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-242024"
 /clone_1fb="RP11-11.1"
 /clone_1fb="RP11-11.1"
 misc_feature
 17380_17440
 /note="Sequence from overlapping clone
 RP11-467D18 (AL590729). Assembly confirmed by restriction
 digest."
 57015_57148
 /note="Sequence from overlapping clone

Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RPS-927M24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RPS-927M24 is at 1 in this sequence. The true right end of clone RPS-21896 is at 6581 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPS-927M24 is from the library RPS-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

FEATURES

source

1. 121922

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/clone="RPS-927M24"

/clone_11b="RPS-5"

1. 248

/note="AluX repeat: matches 59. .306 of consensus"

283. .318

/note="AluU repeat: matches 1. .36 of consensus"

402. .659

/note="AluU repeat: matches 43. .297 of consensus"

1173. 1465

/note="AluX repeat: matches 1. .299 of consensus"

/complement(1611. .2023)

/note="match: GSS: Em:AQ428420"

1808. .1884

/note="MER1C repeat: matches 63. .140 of consensus"

2584. .2777

/note="MLT1A repeat: matches 179. .356 of consensus"

2898. 3524

/note="CPG island"

/evidence="not_experimental"

3023. .7234

/gene="dJ927M24.1"

join(<3023. .3325,3779. .3936,4975. .5043,6649. .6760,6846. .57234)

/gene="dJ927M24.1"

/product="dJ927M24.1.3 (Novel protein, isoform 3)"

/note="match: ESTs: Em:BE747833 Em:BE663360 Em:BE883052 Em:BE277445 Em:BE27383 Em:BE730060 Em:BE395751"

/evidence="not_experimental"

join(<3037. .3325,3779. .3936,4243. .4406,4975. .5043,6649. .6760,6846. .57234)

/gene="dJ927M24.1"

/product="dJ927M24.1.2 (Novel protein, isoform 2)"

/note="match: ESTs: Em:AM375563 Em:R13556 Em:R14470 Em:R87094 Em:AA827892 Em:T91926 Em:H40444

match: proteins: Sw:P22398"

/evidence="not_experimental"

3564. .3690

/note="MIR repeat: matches 108. .238 of consensus"

join(<3637. .3936,4975. .5043,6649. .6760)

/gene="dJ927M24.1"

/product="dJ927M24.1.1 (Novel protein, isoform 1)"

/note="match: ESTs: Em:H26725 Em:H4044 Em:R46422 Em:H22204 Em:A1808326 Em:A115783 Em:R87094 Em:AA464847 Em:A181822 Em:A1003178 Em:AW515615 Em:T91926 Em:A1924555 Em:AW299397 Em:A1276171 Em:AA777670 Em:AW510700 Em:A1084790 Em:AA411324 Em:AW189888 Em:A1760566 Em:A175468 Em:R13596 Em:T25024 Em:AW628424 Em:AA468277 Em:R14470 Em:AW375363 Em:AA468204 Em:AA470331 Em:AA491735 Em:AA468450 Em:AA470347 Em:AA827892"

misc_feature

/evidence="not_experimental"

repeat_region

/gene="dJ927M24.1"

misc_feature

/note="match: STS: Em:HS83J11T"

polyA_signal

/note="2 copies 68 mer 98% conserved"

polyA_site

/note="match: STS: Em:G22570"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

repeat_region

/gene="dJ927M24.1"

Assembly program: Phrap; version 2.0
Quality coverage: 7.96x in Q20 bases; sum-of-contigs

```

/rpc_family="L1"
repeat_region
2507.2724
/rpc_family="L2"
repeat_region
2727.2752
/rpc_family="(TSGA)n"
repeat_region
3288.4371
/rpc_family="L1"
repeat_region
4658.4698
/rpc_family="L2"
repeat_region
4699.4732
/rpc_family="(CA)n"
repeat_region
4733.4823
/rpc_family="L2"
repeat_region
4837.4860
/rpc_family="(TGA)n"
repeat_region
5466.5560
/rpc_family="Alu"
repeat_region
6580.6745
/rpc_family="MIR"
repeat_region
7545.7832
/rpc_family="Alu"
repeat_region
7930.8235
/rpc_family="Alu"
repeat_region
8511.8825
/rpc_family="Alu"
repeat_region
9568.9633
/rpc_family="A-rich"
repeat_region
10074.10967
/rpc_family="L1"
repeat_region
10993.11405
/rpc_family="L1"
repeat_region
11406.11572
/rpc_family="MIR"
repeat_region
11573.11854
/rpc_family="Alu"
repeat_region
11855.12118
/rpc_family="MIR"
repeat_region
12140.13287
/rpc_family="L1"
repeat_region
13331.13580
/rpc_family="L1"
repeat_region
13765.14052
/rpc_family="Alu"
repeat_region
14101.14648
/rpc_family="L1"
repeat_region
14696.14985
/rpc_family="Alu"
repeat_region
15005.15289
/rpc_family="Alu"
repeat_region
16240.16448
/rpc_family="MIR2_type"
repeat_region
16550.16602
/rpc_family="(CA)n"
repeat_region
16964.17021
/rpc_family="Mariner"
repeat_region
18659.18925
/rpc_family="Alu"
repeat_region
19038.19060
/rpc_family="AT-rich"
repeat_region
19108.19169
/rpc_family="L2"
misc_feature
19720.21664
/Note="probable heat shock protein (HSP-70) pseudogene"
repeat_region
22353.22453
/rpc_family="L2"
repeat_region
23480.23603
/rpc_family="MIR"
repeat_region
24035.24266
/rpc_family="L1"
repeat_region
24300.25397
/rpc_family="L1"
repeat_region
25398.25708
/rpc_family="Alu"

```

```

repeat_region 25709.26893
/rpc_family="L1"
repeat_region 26911.27044
/rpc_family="L1"
repeat_region 27097.27126
/rpc_family="AT-rich"
repeat_region 27962.27993
/rpc_family="(CATAA)n"
repeat_region 28726.28829
/rpc_family="(TAAA)n"
repeat_region 30880.30902
/rpc_family="(CAA)n"
repeat_region 31045.31155
/rpc_family="MIR"
repeat_region 31326.31455
/rpc_family="Alu"
repeat_region 31456.31516
/rpc_family="purine-rich"
repeat_region 31524.31628
/rpc_family="(GAAA)n"
repeat_region 31842.31922
/rpc_family="MIR"
repeat_region 32066.32377
/rpc_family="Alu"
repeat_region 34619.34735
/rpc_family="(TA)n"
repeat_region 34736.34759
/rpc_family="(GA)n"
repeat_region 34842.35109
/rpc_family="MIR"
repeat_region 35446.35485
/rpc_family="AT-rich"
repeat_region 35676.35854
/Note="match to EST AA682775 (NID:92669458) ah52h07.s1"
repeat_region 35792.36098
/rpc_family="Alu"
repeat_region 36614.36698
/rpc_family="L2"
repeat_region 38042.38334
/rpc_family="Alu"
misc_feature 38235.38521
/Note="match to EST AA682775 (NID:92669458) ah52h07.s1"
repeat_region 38623.38655
/rpc_family="(CA)n"
repeat_region 39148.39201

```

Query Match 2.8%; Score 71; DB 9; Length 127824;
 Best Local Similarity 100.0%; Pred.No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGACCTGATGATCCACCCACCTTGCCAAAGTCGTGATTCAG 2109
 |||||
 DB 120490 TCGAAGCTCTGACCTGATGATCCACCCACCTTGCCAAAGTCGTGATTCAG 120431
 |||||
 QY 2110 GTGTGAGCCAC 2120
 |||||
 DB 120430 GTGTGAGCCAC 120420

RESULT 37
 AC069548/c 131909 bp DNA linear PRI 09-APR-2002
 LOCUS Homo sapiens chromosome 10 clone RP11-522H2, complete sequence.
 DEFINITION AC069548
 ACCESSION AC069548
 VERSION AC069548.6 GI:20087107
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 131909)
 AUTHORS Smith P.R.
 TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome

Sequence Data

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 131909)
AUTHORS Smith, D.R.

JOURNAL Direct Submission
TITLE Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA

REFERENCE 3 (bases 1 to 131909)
AUTHORS Smith, D.R.

JOURNAL Direct Submission
TITLE Submitted (26-FEB-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA

REFERENCE 4 (bases 1 to 131909)
AUTHORS Smith, D.R.

JOURNAL Direct Submission
TITLE Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA

REFERENCE 5 (bases 1 to 131909)
AUTHORS Smith, D.R.

JOURNAL Direct Submission
TITLE Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA

COMMENT On Apr 9, 2002 this sequence version replaced gi:19774381.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: <http://www.genomecorp.com/>
Contact: gic-seqcenter@genomecorp.com
Project Information
Center project name: hg231

IMPORTANT: This sequence is not the entire
insert of clone Rpl1-522H2. It may be
shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone Rpl1-354E23 is at
129909 in this sequence.

----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 990315

FEATURES
source Location/Qualifiers
1..131909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="Rpl1-522H2"
/clone_1fb="RPC1-11"

BASE COUNT 40940 a 26901 c 26639 g 37429 t
ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 131909;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTGACCTGAGTATCCACCCACCTGCGCTCCCAAGTCTGGATTACG 2109
Db 85510 TCGAAGCTGACCTGAGTATCCACCCACCTGCGCTCCCAAGTCTGGATTACG 85451

Qy 2110 GTGTGAGCCAC 2120
Db 85450 GTGTGAGCCAC 85440

RESULT 38
LOCUS AL589792 132592 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP5-889P23 map p36.21-36.33, ***
ACCESSION AL589792
VERSION AL589792.1 GI:13277194
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.

SOURCE

human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 132592)
AUTHORS Lloyd, C.

JOURNAL Direct Submission
TITLE Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
Requests: clonerequests@sanger.ac.uk

COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: db889P23

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 12% of reads
Sequencing vector: plasmid; L08752; 87% of reads
Chemistry: Dye-terminator ABI; 14% of reads
Chemistry: Dye-terminator Big Dye; 60% of reads
Chemistry: Dye-terminator Big Dye; 24% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Consensus quality: 131458 bases at least Q40
Consensus quality: 131709 bases at least Q30
Consensus quality: 131965 bases at least Q20
Insert size: 132292; sum-of-contigs
Insert size: 132129; 14.2% error; agarose-fp
Quality coverage: 11.37x in Q20 bases; sum-of-contigs Quality
coverage: 11.65x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3331: contig of 3331 bp in length.
* 3332 3431: gap of 100 bp
* 3432 90998: contig of 87567 bp in length
* 90999 91098: gap of 100 bp
* 91099 130106: contig of 39008 bp in length
* 130107 130206: gap of 100 bp
* 130207 132592: contig of 2386 bp in length.

FEATURES
source Location/Qualifiers
1..132592
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.21-36.33"
/clone="RP5-889P23"
/clone_1fb="RPC1-5"

misc_feature 1..3331
/note="assembly_fragment:00799"
misc_feature 3432..90998
/note="assembly_fragment:02114"
misc_feature 91099..130106
/note="assembly_fragment:03033.0"
misc_feature 130207..132592
/note="assembly_fragment:03380"
BASE COUNT 26843 a 36833 c 36775 g 31841 t 300 others
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 132592;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTGACCTGAGTATCCACCCACCTGCGCTCCCAAGTCTGGATTACG 2109
|||||

Db 105811 TCGAATCTGAGCTGAGTGATCCACCCTTGCTCCCAAGTGGGATTACAG 105870
 QY 2110 GTGTGAGCCAC 2120
 |||||
 Db 105871 GTGTGAGCCAC 105881

RESULT 39
 AC007644 141048 bp DNA linear HTG 23-MAY-1999
 AC007644 Homo sapiens chromosome 17 clone hRPK.19 F.16 map 17, ***
 DEFINITION SQUENCING IN PROGRESS *** 17 unordered pieces.
 AC007644
 AC007644
 AC007644.1 GI:4884012
 HTG, HTGS_PHASE1
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 141048)
 Homo sapiens chromosome 17, clone hRPK.19 F.16
 2 (bases 1 to 141048)
 Unpublished
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckwith, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collins, A.,
 Cooke, P., Dearrellano, K., Depayre, E., Devon, C., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, M., Forrest, C.,
 Finkbeiner, R., Gage, D., Galagan, J., Gardina, S., Gilbert, D., Grant, G.,
 Haggis, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lebecky, J., Lien, C., Locke, K., MacDonald, P.,
 Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
 Melidim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Nijhoff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Tesfaye, S., Torrella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 Submitted (23-MAY-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 448: contig of 448 bp in length
 449 2872: contig of 2424 bp in length
 2873 6852: contig of 3980 bp in length
 6853 10142: contig of 3290 bp in length
 10143 14836: contig of 4694 bp in length
 14837 22329: contig of 7493 bp in length
 22330 27992: contig of 5663 bp in length
 27993 33787: contig of 5795 bp in length
 33788 40261: contig of 6474 bp in length
 40262 45996: contig of 5735 bp in length
 gap of unknown length

45997 53175: contig of 7179 bp in length
 53176 59404: contig of 6229 bp in length
 59405 69232: contig of 9828 bp in length
 69233 78639: contig of 9407 bp in length
 78640 94682: contig of 16043 bp in length
 94683 127542: contig of 32860 bp in length
 127543 141048: contig of 13506 bp in length.
 Location/Qualifiers
 source
 1..141048
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="hRPK.19 F.16"
 /clone_11b="RPCT-11 human BAC library"
 BASE COUNT 38166 a 30217 c 31995 g 40108 t 562 others
 ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 141048;
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAATCTGAGCTGAGTGATCCACCCTTGCTCCCAAGTGGGATTACAG 2109
 Db 134616 TCGAATCTGAGCTGAGTGATCCACCCTTGCTCCCAAGTGGGATTACAG 134675
 QY 2110 GTGTGAGCCAC 2120
 |||||
 Db 134676 GTGTGAGCCAC 134686

RESULT 40
 AC068266 149235 bp DNA linear HTG 24-AUG-2002
 AC068266 Homo sapiens chromosome 2 clone RP11-146F22 map 2, WORKING DRAFT
 DEFINITION
 AC068266
 AC068266.2 GI:8225539
 VERSION HTG, HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 149235)
 Homo sapiens chromosome 2, clone RP11-146F22
 2 (bases 1 to 149235)
 Unpublished
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukingalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardina, S., Gande, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Haggis, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazeres, R., Landers, T., Lebecky, J.,
 Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Melidim, J., Menue, L., Mihova, T., Miranda, C., Menga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schaner, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,


```

misc_feature /note="assembly_fragment"
44019..49427
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
49528..55671
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
55772..61066
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
61167..67660
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
67761..73637
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
73738..81886
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
81987..92046
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
92147..105689
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
105790..117433
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
117534..130091

```

```

Query Match 2.8%; Score 71; DB 2; Length 149235;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2050 TCGAACCTCTGACCTGATGATCCACCCCTTGGCTCCCAAGTCTGGGATTACAG 2109
DB 49134 TCGAACCTCTGACCTGATGATCCACCCCTTGGCTCCCAAGTCTGGGATTACAG 49193

```

```

QY 2110 GTGTGAGCCAC 2120
DB 49194 GTGTGAGCCAC 49204

```

```

RESULT 41
HS1177B19 151188 bp DNA linear PRI 23-NOV-1999
LOCUS HS1177B19
DEFINITION Human DNA sequence from clone 1177B19 on chromosome 1p26.12-36.31.
Contains the 3' part of the DNA-binding zinc finger protein RIZ
gene, ESTs, an STS, GSSs and a CpG island, complete sequence.

```

```

ACCESSION AL031277.1 GI:4375907
VERSION HTG; CpG island; RIZ; zinc finger.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.

```

```

REFERENCE 1 (bases 1 to 151188)
AUTHORS Bagguley, C.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

```

```

COMMENT On Mar 7, 1999 this sequence version replaced gi:4153927.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence is the entire insert of clone 1177B19. The true right
end of clone 726F20 (AL031273) is at 37082 in this sequence. This
sequence has been finished according to sequence map criteria as
follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

```

```

This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
1177B19 is from the library RCT15 constructed at the Roswell Park

```

```

FEATURES
source
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: PCYPAC2.
Location/Qualifiers
1..151188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.12-36.31"
/clone="RP5-117B19"
/clone_lib="RPC1-5"
1..316
/note="LIPAL3 repeat: matches 5826..6156 of consensus"
316..387
/note="12 copies 6 mer gaaga 75% conserved"
317..390
/note="37 copies 2 mer ag 76% conserved"
400..450
/note="MIR repeat: matches 204..256 of consensus"
1157..1537
/note="LIR16A repeat: matches 25..440 of consensus"
2143..2937
/note="12 repeat: matches 1824..2712 of consensus"
2946..3418
/note="WTRC repeat: matches 6..466 of consensus"
4592..4898
/note="AluSc repeat: matches 1..307 of consensus"
4917..5014
/note="LIMB8 repeat: matches 6070..6171 of consensus"
5910..6407
/note="AluY repeat: matches 105..302 of consensus"
complement(6172..6532)
/note="match: GSS A0053455"
7283..7573
/note="AluB repeat: matches 1..294 of consensus"
7641..7931
/note="AluX repeat: matches 3..294 of consensus"
8682..8878
/note="MIR repeat: matches 11..247 of consensus"
9026..9042
/note="NAL1 repeat: matches 240..255 of consensus"
9043..9342
/note="AluSq repeat: matches 1..299 of consensus"
9343..9505
/note="HAL1 repeat: matches 255..431 of consensus"
9562..10370
/note="NAL1 repeat: matches 517..1388 of consensus"
10418..10515
/note="MIR2 repeat: matches 2..99 of consensus"
10638..10822
/note="NAL1 repeat: matches 1506..1707 of consensus"
10923..11142
/note="MIR repeat: matches 13..245 of consensus"
10952..11352
/note="match: GSS A000754"
11227..11466
/note="MIR1 repeat: matches 1..245 of consensus"
11493..13283
/note="LIPAL4 repeat: matches 4106..5885 of consensus"
13284..13628
/note="THB1C repeat: matches 1..371 of consensus"
13629..15238
/note="THB1C-internal repeat: matches 3..1580 of
consensus"
15241..15600
/note="THB1C repeat: matches 1..371 of consensus"
15601..15865
/note="LIPAL4 repeat: matches 5881..6149 of consensus"
15922..16083
/note="LIMC/D repeat: matches 5311..5477 of consensus"
16120..16369
/note="LIMC repeat: matches 1756..1714 of consensus"
16370..16885
/note="MTR1D repeat: matches 1..505 of consensus"

```

```

repeat_region 16886..17391
/note="LMBC repeat: matches 1713..2197 of consensus"
repeat_region 17407..17588
/note="LM4 repeat: matches 2433..2617 of consensus"
repeat_region 17593..17920
/note="THEIC repeat: matches 46..371 of consensus"
repeat_region 17921..18278
/note="THEIB repeat: matches 1..364 of consensus"
repeat_region 18280..18589
/note="THEIB-INTERNAL repeat: matches 1269..1579 of consensus"
repeat_region 19059..19232
/note="MER6 repeat: matches -64..838 of consensus"
repeat_region 19233..20511
/note="THEIB-INTERNAL repeat: matches 5..1284 of consensus"
repeat_region 20516..20871
/note="THEIB repeat: matches 1..364 of consensus"
repeat_region 20872..20926
/note="MSTA repeat: matches 1..55 of consensus"
repeat_region 20927..22462
/note="MSTA-internal repeat: matches 1..1580 of consensus"
repeat_region 22463..22839
/note="THEIC repeat: matches 1..371 of consensus"
repeat_region 22840..24495
/note="LM4 repeat: matches 2620..4287 of consensus"
repeat_region 24496..24790
/note="ALUSg repeat: matches 1..292 of consensus"
repeat_region 24791..25222
/note="LM4 repeat: matches 4287..4716 of consensus"
repeat_region 25246..25474
/note="LM4 repeat: matches -1415..-1183 of consensus"
repeat_region 25497..25963
/note="LTR22 repeat: matches 1..505 of consensus"
repeat_region 25974..26086
/note="LIP repeat: matches 1703..1815 of consensus"
repeat_region 26038..27013
/note="LIPAL0 repeat: matches 2626..3615 of consensus"
repeat_region 27014..27082
/note="THEIB repeat: matches 1..67 of consensus"
repeat_region 27122..27413
/note="THEIC repeat: matches 80..371 of consensus"
repeat_region 27436..28294
/note="LIPAL0 repeat: matches 3644..5531 of consensus"
repeat_region 29317..29346
/note="5 copies 6 mer ttaata 100% conserved"
repeat_region 29348..29767
/note="LIPAL4 repeat: matches 5727..6146 of consensus"
repeat_region 30002..30195
/note="MIR repeat: matches 58..261 of consensus"
repeat_region 30337..30974
/note="match: GSS AQ307663"
repeat_region 30976..31286
/note="LIPAL0 repeat: matches 5514..6158 of consensus"
repeat_region 31287..31286
/note="ALUY repeat: matches 1..311 of consensus"
repeat_region 31324..31626
/note="ALUSg repeat: matches 1..308 of consensus"
repeat_region 31743..32886
/note="LIMD1 repeat: matches 4734..5889 of consensus"
repeat_region 32897..33213
/note="ALUB repeat: matches 1..303 of consensus"
repeat_region 33214..33513
/note="LIMD1 repeat: matches 5889..6185 of consensus"
repeat_region 33525..34025
/note="match: GSS AQ332604"
repeat_region 33828..33989
/note="MUT1 repeat: matches 241..409 of consensus"
repeat_region 34185..34479
/note="MUT1B repeat: matches 8..308 of consensus"
repeat_region 35301..35771
/note="MER42 repeat: matches 1..504 of consensus"
repeat_region 36083..36393

```

```

repeat_region 37864..37951
/note="ALUSx repeat: matches 1..308 of consensus"
repeat_region 38243..38397
/note="MIR repeat: matches 84..173 of consensus"
repeat_region 38453..38752
/note="MIR repeat: matches 61..227 of consensus"
repeat_region 39596..39774
/note="ALUSg repeat: matches 1..301 of consensus"
repeat_region 40042..40500
/note="MER53 repeat: matches 1..182 of consensus"
repeat_region 40546..40636
/note="L2 repeat: matches 2150..2711 of consensus"
repeat_region 41856..41997
/note="MER96 repeat: matches 85..175 of consensus"
repeat_region 42015..42168
/note="MER5A repeat: matches 40..186 of consensus"
repeat_region 42015..42168
/note="MIR repeat: matches 34..199 of consensus"

Query Match 2.84; Score 71; DB 9; Length 151188;
Best Local Similarity 100.0%; Pred. No. 1,2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 2050 TCGAAGCTGACCTGATGATCCACCACTGCGCTCCCAAGTCTGGATTACG 2109
Db 38534 TCGAAGCTGACCTGATGATCCACCACTGCGCTCCCAAGTCTGGATTACG 38475

Cy 2110 GTGTGAGCCAC 2120
Db 38474 GTGTGAGCCAC 38464

RESULT 42
AC026110/C
LOCUS
DEFINITION
Homo sapiens 12 BAC RP11-434B3 (Roswell Park Cancer Institute Human
AC026110
VERSION
AC026110.42 GI:14589352
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 153190)
Munry,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albiooke,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbata,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaý,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., Delaney,K.R., Delgado,O.,
Devay-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denny,A.L., Ding,Y., Dinh,H.H., Douhaite,K.J., Dreper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Eberling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
Gunnarson,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hogson,A., Hogue,M., Holloway,C., Hollins,B., Homel,F.,
Howard,S., Huber,J., Huylk,S., Hune,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Liu,C., Liu,J., Liu,W.,
Lewised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martindale,A., Martinez,E., Massey,E., Maxwell,M.P.,
Martindale,A., Martinez,E., Massey,E., Maxwell,M.P.,
Meador,M., Mei,G., Mercher,S., Metzger,M., Miller,A., Miner,G.,
Minter,Z., Mitchell,T., Monabdat,K., Montgomery,K.I., Morgan,W.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokweto,S.,

```



```

repeat_region      /rpt_family="12"
16163.16588
repeat_region      /rpt_family="11P13"
17097.17130
repeat_region      /rpt_family="(TTTA)n"
complement(17131..17423)
repeat_region      /rpt_family="A1u5x"
17915.17946
repeat_region      /rpt_family="(A)n"
18738.18865
repeat_region      /rpt_family="FLAM_A"
complement(19181..19273)
repeat_region      /rpt_family="MIR"
complement(20066..20328)

```

```

Query Match      2.8% Score 71; DB 9; Length 153190;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2050 TCGAAGCTGCTGAGCTGATCCAGCCACCTTGCTCCCAAGTGTGGGATTACAG 2109
Db 132311 TCGAAGCTGCTGAGCTGATCCAGCCACCTTGCTCCCAAGTGTGGGATTACAG 132252

```

```

Qy 2110 GTGTGAGGCAC 2120
Db 132251 GTGTGAGGCAC 132241

```

```

RESULT 43
AL513012/c 153553 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-26802, *** SEQUENCING IN
DEFINITION PROGRESS ***, 26 unordered pieces.
ACCESSION AL513012
VERSION AL513012.4 GI:12733808
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 153553)
McLay, K.
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requesters: clonerequest@sanger.ac.uk
On Feb 9, 2001 this sequence version replaced gi:12584494.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA26802
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator ABI; 1% of reads
Chemistry: Dye-terminator ET-amersham; 8% of reads
Dye-terminator Big Dye; 92% of reads
Consensus quality: 140250 bases at least Q40
Consensus quality: 145215 bases at least Q30
Consensus quality: 148460 bases at least Q20
Insert size: 151053; sum-of-coverage
Insert size: 200511; 5.3% error; agarose-fp
Quality coverage: 5.07x in Q20 bases; sum-of-coverage
Quality coverage: 4.44x in Q20 bases; agarose-fp

```

* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

FEATURES

source

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2079 2178: contig of 2078 bp in length
2179 10085: contig of 7907 bp in length
10086 10185: gap of 100 bp
10186 12349: contig of 2164 bp in length
12350 12449: gap of 100 bp
12450 25415: contig of 12966 bp in length
25416 25519: gap of 104 bp
25520 28046: contig of 2527 bp in length
28047 28146: gap of 100 bp
28147 30184: contig of 2038 bp in length
30185 30284: gap of 100 bp
30285 62042: contig of 31758 bp in length
62043 62142: gap of 100 bp
62143 69938: contig of 7696 bp in length
69939 69938: gap of 100 bp
69939 72212: contig of 2274 bp in length
72213 72312: gap of 100 bp
72313 74404: contig of 2092 bp in length
74405 74504: gap of 100 bp
74505 78593: contig of 4089 bp in length
78594 78693: gap of 100 bp
78694 81831: contig of 3138 bp in length
81832 81931: gap of 100 bp
81932 84050: contig of 2119 bp in length
84051 84150: gap of 100 bp
84151 88043: contig of 3893 bp in length
88044 88143: gap of 100 bp
88144 93493: contig of 5350 bp in length
93494 93593: gap of 100 bp
93594 101833: contig of 8240 bp in length
101834 101933: gap of 100 bp
101934 104335: contig of 2402 bp in length
104336 104335: gap of 100 bp
104336 107388: contig of 2953 bp in length
107389 107488: gap of 100 bp
107489 132346: contig of 24658 bp in length
132347 132446: gap of 100 bp
132447 137068: contig of 4632 bp in length
137069 137168: gap of 100 bp
137169 139335: contig of 2167 bp in length
139336 139435: gap of 100 bp
139436 142713: contig of 3278 bp in length
142714 142813: gap of 100 bp
142814 145741: contig of 2928 bp in length
145742 145841: gap of 100 bp
145842 148055: contig of 2214 bp in length
148056 148155: gap of 100 bp
148156 150735: contig of 2580 bp in length
150736 150835: gap of 100 bp
150836 153553: contig of 2718 bp in length.
Location/Qualifiers
1..153553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-26802"
/clone_id="RP11-1.1"
1..2078
/note="assembly fragment: 00410
fragment chain: 1"
2179..10085
/note="assembly fragment: 00093
fragment chain: 1"
10186..12349
/note="assembly fragment: 01777
fragment chain: 1"
12450..25415
/note="assembly fragment: 00225
fragment chain: 1"

```

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature      25516..28046
                    /note="assembly_fragment:00586
                    fragment_chain:3"
misc_feature      28147..30184
                    /note="assembly_fragment:01302
                    fragment_chain:3"
misc_feature      30285..52042
                    /note="assembly_fragment:00735
                    fragment_chain:3"
misc_feature      62143..69838
                    /note="assembly_fragment:02041
                    fragment_chain:3"
misc_feature      69939..72212
                    /note="assembly_fragment:01438
                    fragment_chain:4"
misc_feature      72313..74404
                    /note="assembly_fragment:02377
                    fragment_chain:4"
misc_feature      74505..78593
                    /note="assembly_fragment:01818
                    fragment_chain:5"
misc_feature      78694..81831
                    /note="assembly_fragment:02340
                    fragment_chain:5"
misc_feature      81932..84050
                    /note="assembly_fragment:02436
                    fragment_chain:6"
misc_feature      84151..88043
                    /note="assembly_fragment:00749
                    fragment_chain:6"
misc_feature      88144..93493
                    /note="assembly_fragment:02596
                    fragment_chain:7"
misc_feature      93594..101833
                    /note="assembly_fragment:02026
                    fragment_chain:7"
misc_feature      101934..104335
                    /note="assembly_fragment:00104"
                    /note="assembly_fragment:00104"
misc_feature      104436..107388
                    /note="assembly_fragment:00544"
                    /note="assembly_fragment:00544"
misc_feature      107489..132346
                    /note="assembly_fragment:00666"
                    /note="assembly_fragment:00666"
misc_feature      132447..137068
                    /note="assembly_fragment:00929"
                    /note="assembly_fragment:00929"
misc_feature      137169..139335
                    /note="assembly_fragment:01276"
                    /note="assembly_fragment:01276"
misc_feature      139436..142713
                    /note="assembly_fragment:01767"
                    /note="assembly_fragment:01767"
misc_feature      142814..145741
                    /note="assembly_fragment:01788"
                    /note="assembly_fragment:01788"
misc_feature      145842..148055
                    /note="assembly_fragment:01844"
                    /note="assembly_fragment:01844"
misc_feature      148156..150735
                    /note="assembly_fragment:02038"
                    /note="assembly_fragment:02038"
misc_feature      150836..153553
                    /note="assembly_fragment:02321"
                    /note="assembly_fragment:02321"
BASE COUNT      39198 a 36089 c 37065 g 38671 t 2530 others
ORIGIN

```

```

Query Match      2.8%: Score 71; DB 2; Length 15353;
Best Local Similarity 100.0%; Pred. No. 1,2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2050 TCGAAGCTCTGAGCTGATTCACACCTGCTGCCAAAGTGTGGATTACAG 2109
      |||
Db 123093 TCGAAGCTCTGAGCTGATTCACACCTGCTGCCAAAGTGTGGATTACAG 123094
      |||
QY 2110 GTGTGAGCCAC 2120
      |||
Db 123093 GTGTGAGCCAC 123023

```

RESULT 44

```

AL445669/c
LOCUS      AL445669      153640 bp      DNA      linear      PRI 27-NOV-2000
DEFINITION Human DNA sequence from clone RP11-163G10 on chromosome 1, complete
sequence.
ACCESSION  AL445669
VERSION    AL445669.9  GI:11414665
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 153640)
REFERENCE  Moore M.
            Direct Submission
            Submitted (25-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: hunquere@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Nov 28, 2000 this sequence version replaced gi:11139975.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unure'
            feature key.

```

COMMENT

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
 http://www.sanger.ac.uk/Project/C.elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1
 RP11-163G10 is from the library RPCT-11.1 constructed at the
 Roswell Park Cancer Institute by the group of Peter de Jong. For
 further details see http://bacpac.med.buffalo.edu/
 VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-163G10.
 Location/Qualifiers
 1..153640
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone_lib="RPCT-11.1"
 /clone_1fb="RPCT-11.1"

FEATURES

source

```

BASE COUNT      46296 a 29313 c 31192 g 46039 t
ORIGIN

```

```

Query Match      2.8%: Score 71; DB 9; Length 153640;
Best Local Similarity 100.0%; Pred. No. 1,2e-26;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2050 TCGAAGCTCTGAGCTGATTCACACCTGCTGCCAAAGTGTGGATTACAG 2109
      |||
Db 144571 TCGAAGCTCTGAGCTGATTCACACCTGCTGCCAAAGTGTGGATTACAG 144512
      |||
QY 2110 GTGTGAGCCAC 2120
      |||
Db 144511 GTGTGAGCCAC 144501

```

```

RESULT 45
AC091181/c      155923 bp      DNA      linear      PRI 18-JUL-2002
LOCUS          Homo sapiens chromosome 17, clone RP11-426D19, complete sequence.
DEFINITION
ACCESSION  AC091181
VERSION    AC091181.8  GI:21903633

```


KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 155923)
Birren, B., Nusbaum, C., and Lander, E.
Homo sapiens chromosome 17, clone RP11-426D19
Unpublished
2 (bases 1 to 155923)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
Camata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M.,
Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K.,
Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W.,
Iliev, I., Johnson, R., Jones, C., Karatas, A., Lacroque, X.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Marquis, N., Mathews, C., McCarthy, M.,
McBwan, P., McKernan, K., McPheeters, R., Meltzer, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Sougnuez, C., Spencer, B., Stange-Thomann, N.,
Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S.,
Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 155923)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C.,
McCarthy, M., Meltzer, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 155923)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C.,
McCarthy, M., Meltzer, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (18-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 18, 2002 this sequence version replaced gi:21699272.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13059
Center clone name: 426_D_19

sfe end overlaps AC015883 [WICGR L725]; T7 end overlaps AC003950
[WICGR project L261]. We are submitting the first 155,9 kilobases
of this project.
----- Location/Qualifiers
1.155923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-426D19"
/clone_lib="RPC1-11 Human Male BAC"
727. 731
/note="<30 qual SNGL region"
repeat_region
991. 1039
/rpt_family="GAAAAA"n"
1463. 1497
/rpt_family="(A)n"
1605. 1609
/note="<30 qual SNGL region"
1662. 1666
/note="<30 qual SNGL region"
1957. 1961
/note="<30 qual SNGL region"
1979. 1983
/note="<30 qual SNGL region"
2263. 2397
/rpt_family="AlusX"
2398. 2706
/rpt_family="AlusG"
2707. 2842
/rpt_family="AlusX"
complement(2861..3165)
/rpt_family="AluY"
3740. 4036
/rpt_family="AluY"
4558. 4666
/rpt_family="WIR3"
complement(6158..6455)
/rpt_family="AlusX"
7397. 7503
/rpt_family="LTR67"
7553. 7723
/rpt_family="WIR3"
complement(7734..7992)
/rpt_family="AlusX"
8405. 8511
/rpt_family="L1ME4A"
8973. 9024
/rpt_family="AT_rich"
9085. 9289
/rpt_family="L1ME4A"
9388. 9408
/rpt_family="(TG)n"
9456. 9581
repeat_region

```

repeat_region      /rpt_family="LIME4"
11142..11299
/rpt_family="LIR"
repeat_region      11869..11938
/rpt_family="MER5B"
12035..12148
/rpt_family="L2"
repeat_region      complement(12263..12382)
/rpt_family="Charliet"
repeat_region      complement(12384..12691)
/rpt_family="AlusX"
repeat_region      complement(12699..12920)
/rpt_family="Charliet"
repeat_region      14673..14742
/rpt_family="LIR"
repeat_region      14748..15036
/rpt_family="AlusX"
repeat_region      complement(15057..15109)
/rpt_family="L2"
repeat_region      15485..15613
/rpt_family="L2"
repeat_region      15730..15798
/rpt_family="LIR3"
repeat_region      complement(16266..16270)
/note="<30 qual SINGL region"
17150..17219
/rpt_family="(TTG)n"
repeat_region      complement(17222..17503)
/rpt_family="AlusX"
repeat_region      complement(17416..17422)
/note="<30 qual SINGL region"
repeat_region      complement(17430..17446)
/note="<30 qual SINGL region"
repeat_region      complement(17468..17472)
/note="<30 qual SINGL region"
repeat_region      complement(17480..17488)
/note="<30 qual SINGL region"
repeat_region      complement(17499..17503)
/note="<30 qual SINGL region"
repeat_region      complement(17646..18071)
/rpt_family="LIMB8"
repeat_region      complement(17979..17983)
/note="<30 qual SINGL region"
repeat_region      complement(18043..18051)
/note="<30 qual SINGL region"
repeat_region      complement(18074..18138)
/note="<30 qual SINGL region"
repeat_region      18788..18791
/note="probably TTTT"
repeat_region      complement(18821..18861)

```

Query Match 2.8%; Score 71; DB 9; Length 155923;
 Best Local Similarity 100.0%; Pred. No. 1.1e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 2050 TCGAAGCTCTGAGCTGATCCACCCCTTGCCCTCCCAAGGCTGGAGATTACG 2109
DB 5894 TCGAAGCTCTGAGCTGATCCACCCCTTGCCCTCCCAAGGCTGGAGATTACG 58895
OY 2110 GTGTGAGCCAC 2120
DB 58884 GTGTGAGCCAC 58874

```

RESULT 46
 AC073839/c
 LOCUS AC073839 158524 bp DNA linear HTG 15-JUL-2000
 DEFINITION Homo sapiens chromosome 17 clone RP11-426D19, WORKING DRAFT
 SEQUENCE 6 unordered pieces.
 ACCESSION AC073839
 VERSION AC073839.1 GI:8810524
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 158524)
 AUTHORS Waterston,R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 158524)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0426D19
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 155268 bases at least Q40
Consensus quality: 15306 bases at least Q30
Consensus quality: 15688 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 158024; sum-of-contigs
Quality coverage: 4.95 in Q20 bases; agarose-fp
Quality coverage: 6.21 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3335 3334: contig of 3334 bp in length
* 3435 3434: gap of unknown length
* 3435 23416: contig of 19982 bp in length
* 23417 23516: gap of unknown length
* 23517 50419: contig of 26903 bp in length
* 50420 50519: gap of unknown length
* 50520 79326: contig of 28807 bp in length
* 79327 79426: gap of unknown length
* 79427 113976: contig of 34350 bp in length
* 113977 114076: gap of unknown length
* 114077 158524: contig of 44448 bp in length.
Location/Qualifiers
1..158524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-426D19"
1..3334
/note="assembly_name:Contig4"
3435..23416
/note="assembly_name:Contig5"
23517..50419
/note="assembly_name:Contig6"
50520..79326
/note="assembly_name:Contig7"
79427..113976
/note="assembly_name:Contig8
clone_end:17
vector_side:left"
114077..158524
/note="assembly_name:Contig9

```


misc_feature /clone="RP11-26P17"
 /clone.lib="RP11-11 Human Male BAC"
 1. .1146
 misc_feature /note="assembly_fragment"
 1247. .2748
 misc_feature /note="assembly_fragment"
 2849. .4365
 misc_feature /note="assembly_fragment"
 4466. .5524
 misc_feature /note="assembly_fragment"
 5625. .7199
 misc_feature /note="assembly_fragment"
 7300. .9128
 misc_feature /note="assembly_fragment"
 9229. 10151
 /note="assembly_fragment"
 /clone_end:T7
 vector_side:right"
 10252. .13233
 misc_feature /note="assembly_fragment"
 13334. .17255
 misc_feature /note="assembly_fragment"
 17356. .19626
 misc_feature /note="assembly_fragment"
 19727. .21287
 /note="assembly_fragment"
 /clone_end:SP6
 vector_side:right"
 21388. .25830
 misc_feature /note="assembly_fragment"
 25931. 30523
 misc_feature /note="assembly_fragment"
 30624. .40384
 misc_feature /note="assembly_fragment"
 40685. .48850
 misc_feature /note="assembly_fragment"
 48951. .54874
 misc_feature /note="assembly_fragment"
 54975. .66089
 misc_feature /note="assembly_fragment"
 66190. .73670
 misc_feature /note="assembly_fragment"
 73771. .83280
 misc_feature /note="assembly_fragment"
 83381. .92734
 misc_feature /note="assembly_fragment"
 92835. 102275
 misc_feature /note="assembly_fragment"
 102376. 114577
 misc_feature /note="assembly_fragment"
 114678. .129572
 misc_feature /note="assembly_fragment"
 129673. .143684
 misc_feature /note="assembly_fragment"
 143785. 161059
 /note="assembly_fragment"

BASE COUNT 42532 a 37330 c 38379 g 40411 t 2407 others
 ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 161059;
 Best Local Similarity 100.0%; Pred. No. 1,1e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2050 TCGAAGCTCGAAGCTGAGTATCCACCACTTGGCTCCCAAGTGTGGATTTCAG 2109
 Db 79231 TCGAAGCTCGAAGCTGAGTATCCACCACTTGGCTCCCAAGTGTGGATTTCAG 79290
 Oy 2110 GTGTGAGCCAC 2120
 Db 79291 GTGTGAGCCAC 79301

RESULT 48

AC025145/c
 LOCUS
 DEFINITION Homo sapiens chromosome 1 clone RP11-26P17 map 1, WORKING DRAFT
 ACCESSION AC025145
 VERSION AC025145.2 GI:7272268
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 1, clone RP11-26P17
 JOURNAL Unpublished
 REFERENCE
 AUTHORS 2 (bases 1 to 161059)
 TITLE
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choquet,Y., Colangelo,M., Collins,S.,
 Collamore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehotzky,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
 Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Miya,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pieter,N.,
 Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talmes,J.,
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 COMMENT
 Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research 370 Charles Street, Cambridge, MA 02141, USA
 On Mar 21, 2000 this sequence version replaced gi:158956.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 Project Information
 Center project name: U7960
 Center clone name: 26_P_17

Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 146317 bases at least Q40
 Consensus quality: 154085 bases at least Q30
 Insert size: 166000; agarose-fp
 Insert size: 158659; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 4.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 25 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 1146: contig of 1146 bp in length

```

* 1147 1246: gap of 100 bp
* 1247 2748: contig of 1502 bp in length
* 2749 2848: gap of 100 bp
* 2849 4365: contig of 1517 bp in length
* 4366 4465: gap of 100 bp
* 4466 5524: contig of 1059 bp in length
* 5525 5624: gap of 100 bp
* 5625 7139: contig of 1575 bp in length
* 7200 7299: gap of 100 bp
* 7300 9128: contig of 1829 bp in length
* 9129 9228: gap of 100 bp
* 9229 10151: contig of 923 bp in length
* 10152 10251: gap of 100 bp
* 10252 13233: contig of 2982 bp in length
* 13234 13333: gap of 100 bp
* 13334 17255: contig of 3922 bp in length
* 17256 17355: gap of 100 bp
* 17356 19626: contig of 2271 bp in length
* 19627 19726: gap of 100 bp
* 19727 21287: contig of 1561 bp in length
* 21288 21387: gap of 100 bp
* 21388 25830: contig of 4443 bp in length
* 25831 25930: gap of 100 bp
* 25931 30523: contig of 4593 bp in length
* 30524 30623: gap of 100 bp
* 30624 40584: contig of 9961 bp in length
* 40585 40684: gap of 100 bp
* 40685 48850: contig of 8166 bp in length
* 48851 48950: gap of 100 bp
* 48951 54874: contig of 5924 bp in length
* 54875 54974: gap of 100 bp
* 54975 66089: contig of 1115 bp in length
* 66090 66189: gap of 100 bp
* 66190 73670: contig of 7481 bp in length
* 73671 73770: gap of 100 bp
* 73771 83280: contig of 9510 bp in length
* 83281 92734: contig of 9354 bp in length
* 92735 92834: gap of 100 bp
* 92835 102275: contig of 9441 bp in length
* 102276 102375: gap of 100 bp
* 102376 114577: contig of 12202 bp in length
* 114578 114677: gap of 100 bp
* 114678 129572: contig of 14895 bp in length
* 129573 129672: gap of 100 bp
* 129673 143684: contig of 14012 bp in length
* 143685 143784: gap of 100 bp
* 143785 161059: contig of 17275 bp in length.

```

FEATURES

```

source
  1.161059
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="1"
    /map="1"
  /clone_1lb="RP11-26P17"
  1.1146
    /note="assembly_fragment"
  /note="assembly_fragment"
  1247.2748
    /note="assembly_fragment"
  2849.4365
    /note="assembly_fragment"
  4466.5524
    /note="assembly_fragment"
  5625.7139
    /note="assembly_fragment"
  7300.9128
    /note="assembly_fragment"
  9229.10151
    /note="assembly_fragment"
  /note="assembly_fragment"
  clone_end:17
  vector_side:right"
misc_feature
  10252..13233

```

```

misc_feature 13334..17255 /note="assembly_fragment"
misc_feature 17356..19626 /note="assembly_fragment"
misc_feature 19727..21287 /note="assembly_fragment"
misc_feature 21388..25830 /note="assembly_fragment"
misc_feature 25931..30523 /note="assembly_fragment"
misc_feature 30624..40584 /note="assembly_fragment"
misc_feature 40685..48850 /note="assembly_fragment"
misc_feature 48951..54874 /note="assembly_fragment"
misc_feature 54975..66089 /note="assembly_fragment"
misc_feature 66190..73670 /note="assembly_fragment"
misc_feature 73771..83280 /note="assembly_fragment"
misc_feature 83381..92734 /note="assembly_fragment"
misc_feature 92835..102275 /note="assembly_fragment"
misc_feature 102376..114577 /note="assembly_fragment"
misc_feature 114678..129572 /note="assembly_fragment"
misc_feature 129673..143684 /note="assembly_fragment"
misc_feature 143785..161059 /note="assembly_fragment"

```

```

BASE COUNT 42532 a 37330 c 38375 g 40411 t 2407 others
ORIGIN

```

Query Match

```

Best Local Similarity 2.84; Score 71; DB 2; Length 161059;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2050 TCGAAGCTCCGACCTCAGGTGATCCACCCACCTGCGCTCCCAAGTGGGATTACG 2109
Db 25024 TCGAAGCTCCGACCTCAGGTGATCCACCCACCTGCGCTCCCAAGTGGGATTACG 24965
Qy 2110 GTGTGAGCCAC 2120
Db 24964 GTGTGAGCCAC 24954

```

```

RESULT 49
AC012664
LOCUS
DEFINITION Homo sapiens clone RP11-3C18, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
ACCESSION AC012664
VERSION AC012664.4 GI:9966277
KEYWORDS HTG: HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 163421)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-3C18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163421)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

```

TITLE
JOURNAL

Cooke, P., Deatrelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galan, J., Gaidys, S., Grant, G., Haggis, B., Hefford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karamas, A., Klein, J.,
Lehoczky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwen, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wymen, D., Ye, M., Zimmer, A. and Zody, M.

Submitted (02-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:1107860.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L2709

Center clone name: 3_C_18

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158604 bases at least Q40

Consensus quality: 161244 bases at least Q30

Consensus quality: 162123 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 162621; sum-of-contigs

Quality coverage: 6.1 in Q20 bases; sum-of-contigs

Quality coverage: 6.2 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3528: contig of 3528 bp in length
* 3529 3628: gap of 100 bp
* 3629 6379: contig of 2751 bp in length
* 6380 6479: gap of 100 bp
* 6480 11952: contig of 5473 bp in length
* 11953 12052: gap of 100 bp
* 12053 87340: contig of 75288 bp in length
* 87341 87440: gap of 100 bp
* 87441 93389: contig of 5949 bp in length
* 93390 93489: gap of 100 bp
* 93490 107761: contig of 14272 bp in length
* 107762 107861: gap of 100 bp
* 107862 125525: contig of 17662 bp in length
* 125526 125625: gap of 100 bp
* 125626 160209: contig of 34584 bp in length
* 160210 160309: gap of 100 bp
* 160310 163421: contig of 3112 bp in length.

FEATURES

source

1. 163421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-3C18"
/clone_id="RP11 Human Male BAC"
1. 3528
misc_feature
/note="assembly_fragment
clone end:SP6
vector_side:left"

misc_feature 3629..6379
/note="assembly_fragment"
misc_feature 6480..11952
/note="assembly_fragment"
misc_feature 12053..87340
/note="assembly_fragment"
misc_feature 87441..93389
/note="assembly_fragment"
misc_feature 93490..107761
/note="assembly_fragment"
misc_feature 107862..125525
/note="assembly_fragment"
misc_feature 125626..160209
/note="assembly_fragment"
misc_feature 160310..163421
/note="assembly_fragment"
clone_end:SP6
vector_side:right

BASE COUNT 47341 a 35177 c 33775 g 46328 t 800 others

ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 163421;
Best local similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2050 TCGAAGCTCTGACTGAGTATCCACCCACTTGGCTTCCCAAGGTCGGGATTACAG 2109

DB 35147 TCGAAGCTCTGACTGAGTATCCACCCACTTGGCTTCCCAAGGTCGGGATTACAG 35206

OY 2110 GTGTGAGCCAC 2120

DB 35207 GTGTGAGCCAC 35217

RESULT 50

LOCUS

AL158137

DEFINITION

Homo sapiens chromosome 6 clone RP1-506112, *** SEQUENCING IN

ACCESSION

AL158137

VERSION

AL158137.4

KEYWORDS

HTG; HTGS_PHASE1; HTGS_CANCELLED.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Almeida, J.

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Dec 17, 2000 this sequence version replaced gi:9931671.

Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

Project Information

Center project name: d4506112

Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 162746 bases at least Q40

Consensus quality: 162931 bases at least Q30

Consensus quality: 163148 bases at least Q20

Insert size: 163516; sum-of-contigs

Insert size: 144035; 25.1% error; agarose-fp

Quality coverage: 7.99% in Q20 bases; sum-of-contigs Quality

coverage: 9.07% in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 25098: contig of 25098 bp in length
 * 25099 25198: gap of 100 bp
 * 25199 100746: contig of 75548 bp in length
 * 100747 100846: gap of 100 bp
 * 100847 146855: contig of 46009 bp in length
 * 146856 146955: gap of 100 bp
 * 146956 163816: contig of 16861 bp in length.

FEATURES

source
 1. 163816
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone_1fb="RP3-506112"
 /clone_1fb="RP3-3"
 1. 25098
 /note="assembly_fragment:01964
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 25199. 100746
 /note="assembly_fragment:00823
 fragment_chain:1"
 100847. 146855
 /note="assembly_fragment:01285
 fragment_chain:1"
 146956. 163816
 /note="assembly_fragment:00518
 clone_end:T7
 vector_side:right"

misc_feature

100847. 146855

misc_feature

100847. 146855

misc_feature

146956. 163816

BASE COUNT 50468 a 34486 c 33411 g 45142 t 309 others
 ORIGIN

Query Match 2.8% Score 71; DB 2; Length 163816;
 Best Local Similarity 100.0%; Pred. No. 1.1e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGAGCTGATCCACCCAGCTGCTCCAAAGTCTGGATTAACAG 2109

DB 24897 TCGAAGCTCTGAGCTGATCCACCCAGCTGCTCCAAAGTCTGGATTAACAG 24956

QY 2110 GTGTGAGCCAC 2120

DB 24957 GTGTGAGCCAC 24967

RESULT 51
 HS120G22 166518 bp DNA linear PRI 21-JUL-2000

LOCUS Human DNA sequence from clone RPI-120G22 on chromosome

DEFINITION 1P36.21-36.33, complete sequence.

ACCESSION AL031847.17 GI:9369286

VERSION HTG.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 166518)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Wray, P.

JOURNAL Direct Submission

COMMENT Submitted (20-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jul 22, 2000 this sequence version replaced gi:9368784.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source database:
 Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Infotration
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1

RP1-120G22 is from the library RP3-1 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone
 RP1-120G22. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true right end of clone RP1-120G22 is at 166518 in this
 sequence. The true left end of clone RP5-889923 is at 90508 in this
 sequence. The true right end of clone RP1-20208 is at 100 in this
 sequence.

FEATURES

source

1. 166518
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p36.21-36.33"
 /clone="RP1-120G22"
 /clone_1fb="RP3-1"
 8814. 9131
 /note="match: GSS: Em:AQ716455"
 8816. 9229
 /note="match: GSS: Em:AQ239786"
 10012. 10541
 /note="match: GSS: Em:AQ678815"
 18389. 18823
 /note="match: GSS: Em:AQ270793"
 25588. 26045
 /note="match: GSS: Em:AQ214655"
 25694. 25967
 /note="match: GSS: Em:AQ097494"
 complement(41087. 41505)
 /note="match: GSS: Em:AQ213363"
 complement(41123. 41505)
 /note="match: GSS: Em:AQ133445"
 45297. 45591
 /note="match: GSS: Em:AF046498"
 49180. 49534
 /note="Single clone region. short insert library only"
 complement(57981. 58764)
 /note="match: GSS: Em:AF158011"
 complement(60582. 61043)
 /note="match: GSS: Em:AQ659221"
 64797. 64841
 /note="Single clone region. short insert library only"
 68250. 68620
 /note="match: STS: Em:G16281"
 70460. 70622
 /note="match: STS: Em:G21080"
 complement(74554. 74895)
 /note="match: GSS: Em:AQ097224"
 complement(79470. 79747)
 /note="match: STS: Em:AF191963"
 complement(79556. 79740)

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature


```

repeat_region /note="MIR repeat: matches 176. .262 of consensus"
20482. .20791 /note="AluX repeat: matches 1. .310 of consensus"
repeat_region 20909. .21207 /note="AluX repeat: matches 1. .299 of consensus"
repeat_region 22005. .22233 /note="AluX repeat: matches 82. .312 of consensus"
repeat_region 22604. .22823 /note="MER2 repeat: matches 122. .345 of consensus"
repeat_region 22824. .23131 /note="AluX repeat: matches 1. .296 of consensus"
repeat_region 23186. .23265 /note="L1ME repeat: matches 5729. .5810 of consensus"
repeat_region 23266. .23575 /note="AluX repeat: matches 1. .301 of consensus"
repeat_region 23576. .23844 /note="L1ME repeat: matches 5470. .5729 of consensus"
repeat_region 24601. .24707 /note="L2 repeat: matches 2600. .2710 of consensus"
repeat_region 26577. .26638 /note="L2 repeat: matches 2437. .2494 of consensus"
repeat_region 26859. .27339 /note="L2 repeat: matches 2303. .2750 of consensus"
repeat_region 27340. .27690 /note="MER1B repeat: matches 1. .364 of consensus"
repeat_region 27691. .28151 /note="L2 repeat: matches 1670. .2303 of consensus"
repeat_region 28855. .28924 /note="MER1I repeat: matches 334. .403 of consensus"
repeat_region 29249. .29459 /note="L2 repeat: matches 1173. .1387 of consensus"
repeat_region 29861. .30006 /note="MER1C repeat: matches 320. .464 of consensus"
repeat_region 30007. .30464 /note="MER1B repeat: matches 69. .519 of consensus"
misc_feature 30934. .31550 /note="match: GSS: Em:AQ374906"
repeat_region 31078. .31167 /note="MER1I repeat: matches 431. .516 of consensus"
repeat_region 31207. .37344 /note="L1PA4 repeat: matches 2. .616 of consensus"
repeat_region 37347. .37784 /note="MER1I repeat: matches 5. .444 of consensus"
repeat_region 37940. .38129 /note="MER1I repeat: matches 251. .409 of consensus"
repeat_region 38667. .38972 /note="AluX repeat: matches 1. .306 of consensus"
misc_feature 39184. .39693 /note="match: GSS: Em:AQ74906"
misc_feature 39197. .39882 /note="match: GSS: Em:AQ581614"
repeat_region 39863. .40454 /note="MER6A repeat: matches 1. .603 of consensus"
repeat_region 40485. .40532 /note="24 copies 2 mer ac 72% conserved"
repeat_region 41968. .42134 /note="L2 repeat: matches 2574. .2750 of consensus"
repeat_region 42133. .42202 /note="MIR repeat: matches 116. .192 of consensus"
repeat_region 42388. .43065 /note="L1M2A repeat: matches 5606. .6292 of consensus"
repeat_region 43143. .43206 /note="AluX/FLAM repeat: matches 2. .65 of consensus"
repeat_region 43379. .43431 /note="MIR repeat: matches 81. .129 of consensus"
repeat_region 43803. .44008 /note="MIR repeat: matches 8. .250 of consensus"
repeat_region 45198. .45503 /note="AluX repeat: matches 2. .303 of consensus"
repeat_region 45571. .45596 /note="13 copies 2 mer ca 96% conserved"
misc_feature complement(45579. .46231)
/note="match: GSS: Em:AQ374882"

```

```

misc_feature complement(45737. .46222)
/note="match: GSS: Em:AQ39959"
repeat_region 46482. .46801 /note="AluX repeat: matches 1. .312 of consensus"
repeat_region 47232. .47523 /note="AluX repeat: matches 1. .312 of consensus"

Query Match 2.8%: Score 71; DB 9; Length 167531;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAATCTGACCTGAGTATCCACCCAGCTTGCCTCCAAAGTCTGGATTACG 2109
Db 108870 TCGAATCTGACCTGAGTATCCACCCAGCTTGCCTCCAAAGTCTGGATTACG 108929

Qy 2110 GTGTGAGCCAC 2120
Db 108930 GTGTGAGCCAC 108940

RESULT 54
AC018636 167592 bp DNA linear PRI 01-MAY-2001
LOCUS Homo sapiens chromosome 7 clone RP11-20F2, complete sequence.
DEFINITION AC018636
ACCESSION AC018636.4 GI:13899420
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 167592)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,
Clendenning, J., Ivey, R.G. and Haugen, E.D.
Direct Submission
Submitted (15-DEC-1999) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 167592)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,
Clendenning, J., Ivey, R.G. and Haugen, E.D.
Direct Submission
Submitted (01-MAY-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On May 1, 2001 this sequence version replaced gi:9719752.

----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu

----- Project Information
Center project name: chr-7
Center clone name: RP11-20F2 (djs173)

----- Summary Statistics
Sequencing vector: plasmid: X52328; 100% of reads
Chemistry: Dye-primer-amersham; 94% of reads
Chemistry: Dye-terminator-Et-amersham; 6% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 285688 bases at least Q40
Consensus quality: 292712 bases at least Q30
Consensus quality: 297476 bases at least Q20
Insert size: 191971; 14.7% error; agarose-fp
Insert size: 169842; sum-of-contigs
Quality coverage: 9.6x in Q20 bases; agarose-fp
Quality coverage: 10.9x in Q20 bases; sum-of-contigs

-----
Overlapping Sequences:
5': RP11-155614 (UWGC:djs356)
3': RP11-78611 (UWGC:djs736)
-----
Sequence Quality Assessment:

```


RESULT 55
AC005073 173592 bp DNA linear HTG 15-JUL-2000
LOCUS Homo sapiens chromosome 7 clone CTB-189J21, WORKING DRAFT SEQUENCE.
DEFINITION 8 unordered pieces.
ACCESSION AC005073 GI:9211524
VERSION AC005073.3 GI:9211524
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 173592)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173592)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 15, 2000 this sequence version replaced gi:7631125.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H RG189J21
Summary Statistics
Sequencing vector: M13: 78k
Sequencing strategy: Dye-terminator Big. Dye; 38k of reads
Chemistry: Dye-terminator Big. Dye; 38k of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 169233 bases at least Q40
Consensus quality: 170233 bases at least Q40
Consensus quality: 170920 bases at least Q20
Insert size: 0000; agarose-fp
Insert size: 172892; sum-of-contigs
Quality coverage: 11.08 in Q20 bases; agarose-fp
Quality coverage: 11.88 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1526: contig of 1526 bp in length
* 1527 1626: gap of unknown length
* 1627 3363: contig of 1737 bp in length
* 3364 3463: gap of unknown length
* 3464 6070: contig of 2607 bp in length
* 6071 6170: gap of unknown length
* 6171 17055: contig of 10885 bp in length
* 17056 17155: gap of unknown length
* 17156 28889: contig of 11734 bp in length
* 28890 28899: gap of unknown length
* 28900 48307: contig of 19318 bp in length
* 48308 48408: gap of unknown length
* 48409 74937: contig of 26530 bp in length
* 74938 75038: gap of unknown length
* 75039 173592: contig of 98555 bp in length.
FEATURES
SOURCE
1. 173592
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="CTB-189J21"

misc_feature 1. 1526
/note="assembly_name:Contig60"
misc_feature 1627. 3363
/note="assembly_name:Contig61"
misc_feature 3464. 6070
/note="assembly_name:Contig62"
misc_feature 6171. 17055
/note="assembly_name:Contig63"
misc_feature 17156. 28889
/note="assembly_name:Contig64"
misc_feature 28890. 48307
/note="assembly_name:Contig65"
misc_feature 48408. 74937
/note="assembly_name:Contig66"
misc_feature 75038. 173592
/note="assembly_name:Contig67"
BASE COUNT 43078 a 42616 c 42747 g 44448 t 703 others
ORIGIN
Query Match 2.8%; Score 71; DB 2; Length 173592;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2050 TCGAAGCTCTGACCTGACGTCACCCACCTTGCCTCCCAAGCGCTGGATTACAG 2109
Db 90661 TCGAAGCTCTGACCTGACGTCACCCACCTTGCCTCCCAAGCGCTGGATTACAG 90720
Qy 2110 GTGTAGGCCAC 2120
Db 90721 GTGTAGGCCAC 90731
RESULT 56
AC073065/c 175345 bp DNA linear PRI 09-JAN-2002
LOCUS AC073065
DEFINITION Homo sapiens BAC clone RP11-14903 from 2, complete sequence.
ACCESSION AC073065
VERSION AC073065.6 GI:15341617
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 175345)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 175345)
AUTHORS VanBrunt,A. and Abbott,A.
TITLE The sequence of Homo sapiens BAC clone RP11-14903
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 175345)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 175345)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 175345)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 29, 2001 this sequence version replaced gi:15022717.

----- Genome Center
Center for Machine Learning

Center: Washington University Genome Sequencing Center
Center code: WUGSC

Center code: WUGSC
Web site: <http://www.wugsc.edu>

Web site: <http://genome.wustl.edu/gsc>
Contact: ganiens@wustl.edu

----- Summary Statistics -----
Contact: baprien@watson.wustl.edu

----- summary statistics -----
Center project name: H NH0149

center project name: n_nh0143003

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Jonn D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPEC1-human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Moon, P. Y., Zhao, B., Frenken, E., Tatenno, M., Caranese, J. J. and de Jong, P. J. (1999). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletzer de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-90U9; the clone sequenced to the right is RP11-122A7. Actual start of this clone is at base position 1 of RP11-14903; actual end is at base position 175345 of RP11-14903.

FEATURES

Data from AC027560 and AC068692 was used to finish this clone AC073065. Polymorphisms have been identified between AC073065 and AC068692.

Source

```

source
location/Qualifiers
1. .175345
/organism="Homo sapiens"
/db xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-14903"
/clone_lib="RPC1-11"
57. .190
/rpc_family="MIR"
532 - 824
/rpc_family="Alu"
905. .1222
/rpc_family="Alu"
1234. .1270
/rpc_family="T-rich"
1242. .1553
/rpc_family="Alu"
1687. .1842
/rpc_family="Alu"
1833. .1893
/rpc_family="AT-rich"
1867. .2288
/rpc_family="U1"
3131. .3159
/rpc_family="U1"

```

repeat_region	/rpt_family="Alu"	3160.3445
repeat_region	/rpt_family="L1"	3446.3933
repeat_region	/rpt_family="AT_rich"	3867.3888
repeat_region	/rpt_family="Alu"	4317.4628
repeat_region	/rpt_family="Alu"	4600.4645
repeat_region	/rpt_family="(A)n"	4677.4698
repeat_region	/rpt_family="AT_rich"	4835.4859
repeat_region	/rpt_family="(T)n"	6479.6675
repeat_region	/rpt_family="L2"	6690.6717
repeat_region	/rpt_family="AT_rich"	6699.6998
repeat_region	/rpt_family="Alu"	7004.7242
repeat_region	/rpt_family="L2"	7241.7282
repeat_region	/rpt_family="(TC)n"	7314.7414
repeat_region	/rpt_family="L2"	7415.7629
misc_feature	/rpt_family="Alu"	7488.7491
repeat_region	/note="match to EST A1735669 (NTD:95057193) at19d06.x1"	7609.7629
repeat_region	/rpt_family="(A)n"	7630.7665
repeat_region	/rpt_family="L2"	8012.8143
repeat_region	/rpt_family="Alu"	8434.8721
repeat_region	/rpt_family="Alu"	8840.8878
repeat_region	/rpt_family="(CAA)n"	9871.9919
repeat_region	/rpt_family="(TTTC)n"	9893.1019
repeat_region	/rpt_family="Alu"	10228.10529
repeat_region	/rpt_family="Alu"	10508.10529
repeat_region	/rpt_family="(A)n"	10941.11293
repeat_region	/rpt_family="MAIR"	11463.11808
repeat_region	/rpt_family="ERV1"	11957.12446
repeat_region	/rpt_family="ERV1"	12571.12657
repeat_region	/rpt_family="MIR"	12657.12691
repeat_region	/rpt_family="(T)n"	12665.12860
repeat_region	/rpt_family="Alu"	13058.13357
repeat_region	/rpt_family="Alu"	13336.13357
repeat_region	/rpt_family="(A)n"	13643.13672
repeat_region	/rpt_family="AT_rich"	13667.13942
repeat_region	/rpt_family="Alu"	14256.14603
repeat_region	/rpt_family="LI"	14370.14393
repeat_region	/rpt_family="AT_rich"	14604.14901

```

repeat_region /rpl_family="Alu"
14902. .16152
/rpl_family="L1"
repeat_region 16154. .16465
/rpl_family="Alu"
repeat_region 16470. .17053
/rpl_family="L1"
repeat_region 17054. .17347
/rpl_family="Alu"
repeat_region 17348. .17572
/rpl_family="L1"
repeat_region 17549. .17660
/rpl_family="L1"
repeat_region 17652. .17676
/rpl_family="(?)n"
repeat_region 17684. .18078
/rpl_family="L1"

```

```

Query Match 2.8%; Score 71; DB 9; Length 175345;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2050 TCGAAGCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGGATTACAG 2109
Db 34099 TCGAAGCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGGATTACAG 34040
OY 2110 GTGTGAGCCAC 2120
Db 34039 GTGTGAGCCAC 34029

```

```

RESULT 57
AL391095 178976 bp DNA linear HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 20 clone RP13-218F6, *** SEQUENCING IN
DEFINITION AL391095
ACCESSION AL391095 GI:9856847
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 178976)
AUTHORS Wilson, S.
TITLES Direct Submission
JOURNAL Submitted (16-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgen@sanger.ac.uk
COMMENT On Aug 18, 2000 this sequence version replaced gi:9798357.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humgen@sanger.ac.uk
----- Project Information
Center project name: DB218F6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: MJ3; M77815; 5% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 17658 bases at least Q40
Consensus quality: 17603 bases at least Q30
Insert size: 178876; sum-of-contigs
Insert size: 184221; 3.3% error; agarose-fp
Quality coverage: 13.33x in Q20 bases; sum-of-contigs Quality
Coverage: 12.95x in Q20 bases; agarose-fp

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 112029; contig of 112029 bp in length
* 112030 112129; gap of 100 bp
* 112130 178976; contig of 66847 bp in length.
Location/Qualifiers
source
1. 178976
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP13-218F6"
/clone_id="RP13-1"
1. 112029
/note="assembly_fragment:00848
clone_end:SP6
vector_side:left"
misc_feature 112130. .178976
/note="assembly_fragment:01607"
BASE COUNT 47478 a 41110 c 40786 g 49501 t 101 others
ORIGIN

```

```

Query Match 2.8%; Score 71; DB 2; Length 178976;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2050 TCGAAGCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGGATTACAG 2109
Db 160413 TCGAAGCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGGATTACAG 160472
OY 2110 GTGTGAGCCAC 2120
Db 160473 GTGTGAGCCAC 160483

```

```

RESULT 58
AC103753 179822 bp DNA linear PRI 25-JUL-2002
LOCUS Homo sapiens chromosome 15, clone RP11-460N1, complete sequence.
DEFINITION AC103753
ACCESSION AC103753 GI:21955078
VERSION HTG:
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 179822)
AUTHORS Birren, B., Nusbaum, C., and Lander, E.
TITLES Unpublished
JOURNAL 2 (bases 1 to 179822)
REFERENCE Birren, B., Linon, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Bogunlavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chapel, I., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Darcianno, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gerdy, N.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamet, A., Karatas, A., Kells, C., Laroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McKean, P., McKenna, K., McPheters, R., Meldrum, J.,
Menais, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Plunkharg, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schipchack, R.,
Seaman, S., Sevaly, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Struss, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,

```



```

repeat_region complement (28074, .28170)
                /rpt_family="AluYb"
repeat_region 28172, .28379
                /rpt_family="L1MC4a"
repeat_region complement (28764, .28873)
                /rpt_family="MIR"
repeat_region complement (28874, .29160)
                /rpt_family="AluYb"
repeat_region complement (29161, .29314)
                /rpt_family="MIR"
repeat_region complement (29495, .29716)
                /rpt_family="MIR"
repeat_region 30060, .30374
                /rpt_family="AluYb8"
repeat_region complement (30492, .30786)
                /rpt_family="AluYb"
repeat_region complement (31594, .31917)
                /rpt_family="W618"
repeat_region complement (31935, .32227)
                /rpt_family="L1ME"

```

```

Query Match      2.8%; Score 71; DB 9; Length 179822;
Best Local Similarity 100.0%; Pred. No. 1,1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2050 TCGAACTCCTGACCTGAGTATCCACCACTTGCTCCCAAGTGTGGATTACAG 2109
      |||
DB 129966 TCGAACTCCTGACCTGAGTATCCACCACTTGCTCCCAAGTGTGGATTACAG 130025
      |||
OY 2110 GTGTGAGCCAC 2120
      |||
DB 130026 GTGTGAGCCAC 130036
      |||

```

```

RESULT 59
AC024036/c 181485 bp DNA linear HTG 20-APR-2000
LOCUS Homo sapiens chromosome X clone RP11-341b6, *** SEQUENCING IN
DEFINITION
AC024036 AC024036.2 GI:7622514
VERSION HTG; HTGS_PHASE1.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 181485)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181485)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Apr 20, 2000 this sequence version replaced gi:7008943.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0341P06
----- Summary Statistics -----
Sequencing vector: pLamB1, 100%
Chemistry: Dye-Primer ET, 100% of reads
Chemistry: Dye-Terminator Big Dye, 0% of reads
Assembly program: Phrap; version 0.990119
Consensus quality: 156467 bases at least Q40
Consensus quality: 165014 bases at least Q30
Consensus quality: 168940 bases at least Q20

```

```

Insert size: 195000; agarose-fp
Insert size: 178185; sum-of-contigs
Quality coverage: 3.16 in Q20 bases; agarose-fp
Quality coverage: 3.58 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 34 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
1138: contig of 1138 bp in length
1238: gap of unknown length
1239: contig of 1424 bp in length
2663: gap of unknown length
2763: contig of 1146 bp in length
3908: gap of unknown length
3909: gap of 1256 bp in length
4009: gap of unknown length
5265: gap of unknown length
5365: contig of 1774 bp in length
7139: gap of unknown length
7239: contig of 2061 bp in length
9399: gap of unknown length
9400: contig of 1921 bp in length
11321: gap of unknown length
11421: contig of 1311 bp in length
12732: gap of unknown length
12832: contig of 2284 bp in length
15116: gap of unknown length
15215: contig of 1759 bp in length
15216: gap of unknown length
16975: contig of 2374 bp in length
17075: gap of unknown length
19449: gap of unknown length
19548: contig of 2642 bp in length
19549: gap of unknown length
22191: gap of unknown length
22290: contig of 2903 bp in length
25194: gap of unknown length
25294: contig of 3219 bp in length
28513: gap of unknown length
28613: contig of 2575 bp in length
31188: gap of unknown length
31288: contig of 3896 bp in length
35183: gap of unknown length
35283: gap of 2845 bp in length
38128: contig of 3947 bp in length
38229: gap of unknown length
42175: contig of 3947 bp in length
42176: gap of unknown length
42275: gap of 3913 bp in length
42276: gap of unknown length
46188: gap of unknown length
46189: gap of 4199 bp in length
46289: gap of unknown length
50486: gap of unknown length
50586: contig of 4549 bp in length
55135: gap of unknown length
55235: gap of unknown length
59426: contig of 4191 bp in length
59427: gap of unknown length
59526: gap of unknown length
66329: contig of 6803 bp in length
66429: gap of unknown length
66430: contig of 5425 bp in length
71854: gap of unknown length
71954: gap of 4222 bp in length
71955: contig of 4222 bp in length
76177: gap of unknown length
76277: contig of 7107 bp in length
83483: gap of unknown length
83484: contig of 6570 bp in length
90053: gap of unknown length
90153: gap of 6442 bp in length
96595: contig of unknown length
96596: gap of unknown length
104875: contig of 8180 bp in length
104975: gap of unknown length
115567: contig of 10552 bp in length
115568: gap of unknown length

```

* 115668 124344: contig of 8677 bp in length
 * 124345 124444: gap of unknown length
 * 124445 133574: contig of 9130 bp in length
 * 133575 133674: gap of unknown length
 * 133675 148741: contig of 15067 bp in length
 * 148742 148841: gap of unknown length
 * 148842 181485: contig of 32644 bp in length.

FEATURES
 source
 1. 181485
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="RP11-341P6"

BASE COUNT 48948 a 39001 c 39658 g 50566 t 3312 others
 ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 181485;
 Best Local Similarity 100.0%; Pred. No. 1,1e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGACCTGAGTGATCCACCACCTTGCTCCCAAGTCTGGATTACAG 2109
 Db 143991 TCGAAGCTCTGACCTGAGTGATCCACCACCTTGCTCCCAAGTCTGGATTACAG 143932

QY 2110 GTGTAGCCAC 2120
 Db 143931 GTGTAGCCAC 143921

RESULT 60
 CNS01DW2 181532 bp DNA linear PRI 20-AUG-2001
 LOCUS Human chromosome 14 DNA sequence BAC R-134E15 of library RPEC1-11
 DEFINITION from chromosome 14 of Homo sapiens (human), complete sequence.
 AL136298 AL136298.5 GI:15282094
 VERSION HTG; HTGS_ACTIVEFIN.
 KEYWORDS HTG; HTGS_ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Heilig, R., Petit, J. L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brocote, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F., Levy, M., Eckenberg, R., Bruls, T., Debardinis, V., Cruaud, C., Gyday, G., Saurin, W. and Weissenbach, J.
 TITLE Sequencing of the human chromosome 14
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 181532)
 GENOSCOPE Genoscope.
 DIRECT SUBMISSION Direct Submission
 SUBMITTED (03-AUG-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 Web : www.genoscope.cns.fr
 On Aug 23, 2001 this sequence version replaced gi:13016586.

COMMENT
 TITLE Genoscope / Centre National de Sequencage
 JOURNAL Center code: GS
 AUTHORS Web site: http://www.genoscope.cns.fr/
 JOURNAL Contact: Seqref@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-320M16 (AC-AL132988)
 Downstream BAC (overlapping the SP6 end) : R-431H16
 Summary Statistics
 Assembly program: Phrap; version 2.0
 Quality coverage: 7.89x in Q20 bases; sum-of-contigs

 Overall quality chart :
 Range : bases
 0
 1 - 9 :

FEATURES
 source
 1. 181532
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="R-134E15"
 /clone_id="RPEC1-11"
 /note="matching EMBL:Z39253
 Rdb:RH44252
 dbSTS:STS25945
 identified using the e-PCR software (G. Schuler)"

Percentage of bases with a quality value >= 40 : 99 %
 Location/Qualifiers
 1. 181532
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="R-134E15"
 /clone_id="RPEC1-11"
 /note="matching EMBL:Z39253
 Rdb:RH44252
 dbSTS:STS25945
 identified using the e-PCR software (G. Schuler)"

BASE COUNT 57126 a 37894 c 35296 g 51216 t
 ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 181532;
 Best Local Similarity 100.0%; Pred. No. 1,1e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGACCTGAGTGATCCACCACCTTGCTCCCAAGTCTGGATTACAG 2109
 Db 103688 TCGAAGCTCTGACCTGAGTGATCCACCACCTTGCTCCCAAGTCTGGATTACAG 103747

QY 2110 GTGTAGCCAC 2120
 Db 103748 GTGTAGCCAC 103758

RESULT 61
 AC010204 183113 bp DNA linear HTG 27-JUN-2001
 LOCUS Homo sapiens chromosome 12 clone RP11-66N16, WORKING DRAFT
 DEFINITION
 SEQUENCE 3 unordered pieces.
 AC010204
 VERSION AC010204.19 GI:14547731
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 183113)
 Mzury, D. M., Adams, C., Adio-Oduola, B., Ali-osman, F. R., Allen, C., Albrooks, S. L., Amaratunga, H. C., Are, J. R., Banks, T., Barbata, J., Benton, U., Blum, K., Blankenburg, K., Bonin, D., Bouck, U., Bowie, S., Brice, M., Brown, E., Brown, M., Bryant, N. P., Buhey, C., Butch, P., Burkett, C., Butrell, K. L., Byrd, N. C., Cartron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J., Barnhart, C., Edgar, D., Edwards, C. C., Elhad, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Hulys, S., Hunne, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryan, J., Koval, C., Kravovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, J. C.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, D., Liu, W.,
 Looisseghe, H., Loozard, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mathiney, E., Mleod, M.P., Moadore, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, S.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nookankoo, S.,
 Ogih, M., Okumura, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
 Ruliz, M., Ren, Y., Rivers, M., Rojas, A., Rojchokan, I., Rolfe, M.,
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshkhat, N.,
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
 Stone, I., Sutcliffe, A., Svatek, A., Taber, P., Tamerisa, K.
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Teliford, B., Thomas, R.,
 Thomas, S., Usman, K., Vasequez, L., Vera, V., Villalón, D., Vinsom, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Wellington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
 Weinsteck, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 183113)
 Morley, K.C.

Direct Submission
 Submitted (15-SEP-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 23, 2000 this sequence version replaced gi:14150373.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMKG
Center clone name: RP11-66N16
----- Summary Statistics
Sequencing vector: Plasmid: M77789
Sequencing vector: M13: L08821
Chemistry: Dye-Primer Bodyfy: 46% of reads
Chemistry: Dye-Terminator Big Dye: 54% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 185911 bases at least Q40
Consensus quality: 190254 bases at least Q20
Consensus quality: 192220 bases at least Q20
Estimated insert size: 185399, sum-of-contigs estimation
Estimated insert size: 25312, agarose-gel estimation
Quality coverage: 5.8x in Q20 bases; agarose-gel estimation
Quality coverage: 7.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_drafc_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1 88408: contig of 88408 bp in length
* 88409 88508: gap of unknown length
* 88509 145920: contig of 57312 bp in length
* 145821 145920: gap of unknown length
* 145921 183133: contig of 37193 bp in length.
Location/Qualifiers
1 .183133

```

BASE COUNT	ORIGIN
56572 a	37358 g 52578 t
37358 g	52578 t
52578 t	200 other

Query Match	2.8%;	Score 71;	DB 2;	Length 183113;	
Best Local Similarity	100.0%;	Pred. No. 1,1e-28;			
Matches	71;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Oy	2050	TCGAACCTTCGACCTCCAGTCGATCCACCCACTTGGCGCTCCGAAAGTCGGGATTACAG	2109		
Db	93651	TCGAACCTTCGACCTCCAGTCGATCCACCCACTTGGCGCTCCGAAAGTCGGGATTACAG	93592		
Oy	2110	GTGTGAGCCAC	2120		
Db	93591	GTGTGAGCCAC	93581		
RESULT 62					
AC002042		185994 bp	DNA	linear	HTG 23-NOV-1999
LOCUS					
DEFINITION	AC002042				
Accession	AC002042				
VERSION	AC002042.1	GI:3075381			
KEYWORDS	HTG; HTGS; PHASE1.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 185994)				
	Loftus,B.J., Kim,U.J., Sneedon,V.P., Kalush,F., Brandon,R., Fuhmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L., Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S., Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.				
TITLE	Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q				
JOURNAL	Genomics 60 (3), 285-308 (1999)				
MEDLINE	99425270				
PubMed	10493829				
REFERENCE	2 (bases 1 to 185994)				
AUTHORS	Adams,M.D., Loftus,B.J., Zhou,L., Phillips,C., Brandon,R.C., Fuhmann,J., Kim,U.J., Kerlavage,A.R. and Venter,J.C.				
TITLE	Human chromosome 16p13 BAC clone C1T987SK-A-180G2				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 185994)				
AUTHORS	Adams,M.D. and Loftus,B.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA				
COMMENT	On Apr 22, 1998 this sequence version replaced gi:3068565. * NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.				
	1	2198: contig of 2198 bp in length			
	2199	2248: gap of unknown length			
	2249	46288: contig of 44040 bp in length			
	46289	46338: gap of unknown length			
	46339	110919: contig of 64581 bp in length			
	110920	110965: gap of unknown length			
	110970	145325: contig of 34256 bp in length			
	145226	145275: gap of unknown length			
	145276	185994: contig of 40719 bp in length.			
FEATURES					
source	1. 185994				
	Location/Qualifiers				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="NA-180G2"				
BASE COUNT	51275 a	43334 c	43047 g	48137 t	201 others
ORIGIN					
Query Match	2.8%;	Score 71;	DB 2;	Length 185994;	

Query Match	2.8%	Score 71	DB 2	Length 185994
-------------	------	----------	------	---------------

Best Local Similarity 100.0%; Pred. No. 1,1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTCTGACCTGAGTATCCACCCAGTGGCTCCCAAGTGGGATTACAG 2109
Db 100564 TCGAAGCTCTGACCTGAGTATCCACCCAGTGGCTCCCAAGTGGGATTACAG 100623

Qy 2110 GTGTAGGCAC 2120
Db 100624 GTGTAGGCAC 100634

RESULT 63
AC022858 187340 bp DNA linear PRI 01-DEC-2001
LOCUS Homo sapiens chromosome 8, clone RP11-326E22, complete sequence.
DEFINITION AC022858
AC022858 GI:17223361
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 187340)
Homo sapiens chromosome 8, clone RP11-326E22
2 (bases 1 to 187340)
Unpublished

Bitren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckert, J., Beda, F.,
Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A.,
Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domingo, M., Doyle, M., Fenebor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,
Landers, T., Lechoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGuirk, A., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Plesni, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N.,
Stojanovic, N., Sudramanian, A., Talamas, D., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A., and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 187340)

Bitren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Baetien, V., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kelle, C., Labèque, K.,
Lamazares, R., Landers, T., Lechoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Plesni, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnepf, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Sudramanian, A., Talamas, D., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Trivis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

TITLE
JOURNAL
Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome

COMMENT
On Dec 1, 2001 this sequence version replaced g1:16647119.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L6184
Center clone name: 326_E22

FEATURES
source
Location/Qualifiers
1..187340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-326E22"
/clone_id="RP11-326E22 Human Male BAC"

repeat_region
23..175
/rpt_family="WIR"
repeat_region
345..375
/rpt_family="(A)n"
complement(840..963)
repeat_region
/rpt_family="WIR"
complement(1341..1434)
repeat_region
/rpt_family="L2"
complement(1946..2122)
repeat_region
2384..2465
/rpt_family="L2"
repeat_region
3276..3387
/rpt_family="WIR3"
repeat_region
/rpt_family="L1ME4A"
complement(10792..11244)
repeat_region
/rpt_family="WERS0"
complement(11297..11439)
repeat_region
12058..12098
/rpt_family="WERS0"
repeat_region
12159..12487
/rpt_family="L2"
repeat_region
12159..12487
/rpt_family="L2"
complement(15060..15129)
repeat_region
16515..16616
/rpt_family="WIR"
complement(16737..17033)
repeat_region
/rpt_family="WLR1K"
complement(18280..18468)
repeat_region
19025..19172
/rpt_family="WIR"
repeat_region
/rpt_family="L2"
complement(19958..20241)
repeat_region
/rpt_family="ALUSX"
complement(122187..22334)
repeat_region
22837..23077
/rpt_family="WLR11"
repeat_region
/rpt_family="WIR"
complement(24301..24436)
repeat_region
/rpt_family="L1PA6"
complement(24591..24889)
repeat_region
25986..26278
/rpt_family="ALUSX"
repeat_region
/rpt_family="ALUBD"
complement(26283..26493)
repeat_region
26528..26814
/rpt_family="WIR"
repeat_region
/rpt_family="ALUBD"
complement(28337..28625)
repeat_region
29450..29765
/rpt_family="L1MC4a"

```

repeat_region /rpt_family="L2"
                complement(29775, .29967)
                /rpt_family="MIR3"
repeat_region 32167, .32243
                /rpt_family="L2"
                complement(36701, .37208)
                /rpt_family="LTR9"
repeat_region complement(37274, .37592)
                /rpt_family="L2"
                complement(37649, .37941)
                /rpt_family="L2"
repeat_region 38084, .38160
                /rpt_family="L2"
                complement(38559, .38559)
                /rpt_family="MIR"
repeat_region 38161, .38559
                /rpt_family="TIGser2a"
                complement(39270, .39382)
                /rpt_family="MIR"
repeat_region complement(39270, .39382)
                /rpt_family="MIR"
                complement(40652, .40773)
                /rpt_family="MIR5B"
repeat_region 41260, .41923
                /rpt_family="L2"
                complement(42777, .42809)
                /rpt_family="(TG)n"
                complement(44555, .44610)
                /rpt_family="Char11e8"
repeat_region complement(44785, .44914)
                /rpt_family="Char11e8"
                complement(45202, .45462)
                /rpt_family="AluSg1"
repeat_region 45541, .45579
                /rpt_family="AT_rich"
                complement(45890, .45890)
                /rpt_family="LIMB4"
repeat_region complement(46380, .46650)
                /rpt_family="LIME"
                complement(46687, .46775)
                /rpt_family="MER5B"
repeat_region 46845, .46967
                /rpt_family="LIM4"
                complement(46993, .47087)
                /rpt_family="GA-rich"
                complement(47110, .47923)
                /rpt_family="LIM4"
repeat_region 48311, .48339
                /rpt_family="AT_rich"
                complement(48340, .48631)
                /rpt_family="AluSg4"
                complement(49127, .49281)
                /rpt_family="LIMB4"
repeat_region complement(49408, .49458)
                /rpt_family="L2"
                complement(49475, .49705)
                /rpt_family="MIR"
repeat_region complement(49723, .49935)
                /rpt_family="MIR"
                complement(49981, .50452)
                /rpt_family="LIM4"
repeat_region 50521, .50633
                /rpt_family="L2"
                complement(51723, .52024)
                /rpt_family="AluY"
repeat_region 52263, .52604
                /rpt_family="MER77"
                complement(53510, .53709)
                /rpt_family="L2"
                complement(53994, .54109)

```

Query Match 2.8%; Score 71; DB 9; Length 187340;
 Best Local Similarity 100.0%; Pred. No. 1,1e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 2050 TCGAAGCTCTCACTCAGTGATCCACCCACTTGGCTCCCAAGTCTGGATTACG 2109
      |||
Db 165831 TCGAAGCTCTCACTCAGTGATCCACCCACTTGGCTCCCAAGTCTGGATTACG 165890
      |||
Oy 2110 GTGTGAGCCAC 2120
      |||
Db 165891 GTGTGAGCCAC 165901

```

```

RESULT 64
AC074234
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-251N10, WORKING DRAFT
SEQUENCE, 46 unordered pieces.
ACCESSION
AC074234
AC074234.5 GI:9958207
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Waterston,R.H.
1 (bases 1 to 187782)
REFERENCE
AUTHORS
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187782)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Sep 1, 2000 this sequence version replaced gi:9937807.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0251N10
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 159363 bases at least Q40
Consensus quality: 16681 bases at least Q30
Consensus quality: 170548 bases at least Q20
Insert size: 18300; agarose-fp
Insert size: 181643; sum-of-ctnigs
Quality coverage: 3.33 in Q20 bases; agarose-fp
Quality coverage: 3.50 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1609: contig of 1609 bp in length
* 1710: gap of unknown length
* 1710 3287: contig of 1578 bp in length
* 3288 3387: gap of unknown length
* 3388 5026: contig of 1639 bp in length
* 5027 5126: gap of unknown length
* 5127 6828: contig of 1702 bp in length
* 6829 6929: gap of unknown length
* 6930 8738: contig of 1710 bp in length
* 8739 9929: contig of 1191 bp in length
* 10029: gap of unknown length

```

10030 11667 contig of 1638 bp in length
11668 11667 gap of unknown length
11768 12894 contig of 1227 bp in length
12995 13094 gap of unknown length
13095 15391 contig of 2297 bp in length
15392 15491 gap of unknown length
15492 17549 contig of 2058 bp in length
17550 17549 gap of unknown length
17550 19100 contig of 1451 bp in length
19101 19200 gap of unknown length
19201 20950 contig of 1750 bp in length
20951 21050 gap of unknown length
21051 22597 contig of 1547 bp in length
22598 22697 gap of unknown length
22698 24935 contig of 2238 bp in length
24936 25035 gap of unknown length
25036 27148 contig of 2113 bp in length
27149 27248 gap of unknown length
27249 29295 contig of 2047 bp in length
29296 29395 gap of unknown length
29396 32904 contig of 3509 bp in length
32905 33004 gap of unknown length
33005 34986 contig of 1982 bp in length
34987 35086 gap of unknown length
35087 38007 contig of 2921 bp in length
38008 38107 gap of unknown length
38108 40741 contig of 2634 bp in length
40742 40841 gap of unknown length
40842 44444 contig of 3603 bp in length
44445 44544 gap of unknown length
44545 48181 contig of 3637 bp in length
48182 48281 gap of unknown length
48282 51543 contig of 3282 bp in length
51544 51643 gap of unknown length
51644 55269 contig of 3626 bp in length
55270 55370 gap of unknown length
55370 58678 contig of 3309 bp in length
58679 58778 gap of unknown length
58779 62391 contig of 3613 bp in length
62392 62491 gap of unknown length
62492 65538 contig of 3047 bp in length
65539 65638 gap of unknown length
65639 69308 contig of 3670 bp in length
69309 69408 gap of unknown length
69409 73476 contig of 4068 bp in length
73477 73576 gap of unknown length
73577 78759 contig of 5183 bp in length
78760 78859 gap of unknown length
78860 83059 contig of 4200 bp in length
83060 83159 gap of unknown length
83160 87616 contig of 4457 bp in length
87617 87716 gap of unknown length
87717 92361 contig of 4645 bp in length
92362 92461 gap of unknown length
92462 97848 contig of 5387 bp in length
97849 97948 gap of unknown length
97949 103924 contig of 5976 bp in length
103925 104024 gap of unknown length
104025 110342 contig of 6318 bp in length
110343 110442 gap of unknown length
110443 116107 contig of 5665 bp in length
116108 116207 gap of unknown length
116208 122197 contig of 5990 bp in length
122198 122297 gap of unknown length
122299 127047 contig of 4750 bp in length
127048 127147 gap of unknown length
127149 133610 contig of 6463 bp in length
133611 133710 gap of unknown length
133711 141207 contig of 7497 bp in length
141208 141307 gap of unknown length
141308 148738 contig of 7431 bp in length
148739 148838 gap of unknown length
148839 156469 contig of 7631 bp in length

FEATURES

Source

* 156470 156569: gap of unknown length
* 156570 165115: contig of 8546 bp in length
* 165116 165215: gap of unknown length
* 165216 176448: contig of 11233 bp in length
* 176449 176548: gap of unknown length
* 176549 187782: contig of 11234 bp in length.
Location/Qualifiers
1..187782
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-251N10"
1..1609
/note="assembly_name:Contig17"
1710..3287
/note="assembly_name:Contig18"
3388..5026
/note="assembly_name:Contig19"
5127..6828
/note="assembly_name:Contig20"
6929..8638
/note="assembly_name:Contig22"
8739..9929
/note="assembly_name:Contig24"
10030..11667
/note="assembly_name:Contig25"
11768..12994
/note="assembly_name:Contig26"
13095..15391
/note="assembly_name:Contig27"
15492..17549
/note="assembly_name:Contig28"
17650..19100
/note="assembly_name:Contig29"
19201..20950
/note="assembly_name:Contig30"
21051..22597
/note="assembly_name:Contig31"
22698..24935
/note="assembly_name:Contig32"
25036..27148
/note="assembly_name:Contig33"
27249..29295
/note="assembly_name:Contig34"
29396..32904
/note="assembly_name:Contig35"
33005..34986
/note="assembly_name:Contig36"
35087..38007
/note="assembly_name:Contig37"
38108..40741
/note="assembly_name:Contig38"
40842..44444
/note="assembly_name:Contig39"
44545..48181
/note="assembly_name:Contig40"
48282..51543
/note="assembly_name:Contig42"
51644..55269
/note="assembly_name:Contig42"
55370..58678
/note="assembly_name:Contig43"
58779..62391
/note="assembly_name:Contig44"
62492..65538

Query Match

2.8%; Score 71; DB 2; Length 187782;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGACATCCGACCCAGGATGATCCACCCAGCTGGCCTCCCAAAGTGTGGATTACAG 2109
DB 28365 TCGACTCTGACCTCAGGATGATCCACCCAGCTTGGCTCCCAAAGTGTGGATTACAG 28424


```

repeat_region /rpt_family="(CAT)n"
27405..27462
repeat_region /rpt_family="MIR"
28647..28780
repeat_region /rpt_family="MIR"
28974..29307
repeat_region /rpt_family="Alu"
29347..29377
repeat_region /rpt_family="AT-rich"
29387..29682
repeat_region /rpt_family="Alu"
29905..30445
repeat_region /rpt_family="ERV1"
30462..30753
repeat_region /rpt_family="Alu"
30776..30800
repeat_region /rpt_family="AT-rich"
31605..32322
repeat_region /rpt_family="L2"
32519..32577
repeat_region /rpt_family="MER103"
34229..34657
repeat_region /rpt_family="MALR"
34658..34988
repeat_region /rpt_family="Alu"
34985..35051
repeat_region /rpt_family="MALR"
39145..39166
repeat_region /rpt_family="AT-rich"
41247..41546
repeat_region /rpt_family="Alu"
41769..41854
repeat_region /rpt_family="CR1"
41861..41917
repeat_region /rpt_family="CR1"
41930..42101
repeat_region /rpt_family="Alu"
42102..42124
repeat_region /rpt_family="(TTTGG)n"
42125..42244
repeat_region /rpt_family="Alu"
42671..42729
repeat_region /rpt_family="(TA)n"
43160..43230
repeat_region /rpt_family="MIR"
43558..43945
repeat_region /rpt_family="L2"
44048..44458
repeat_region /rpt_family="L2"
44604..44722
repeat_region /rpt_family="MIR"
50651..50957
repeat_region /rpt_family="Alu"
56138..56203
repeat_region /rpt_family="(TAGA)n"
56249..56331
repeat_region /rpt_family="GA-rich"
56341..56518
repeat_region /rpt_family="(GGA)n"
56553..56657
repeat_region /rpt_family="GA-rich"
57763..57871

```

Query Match 2.8%; Score 71; DB 9; Length 188829;
 Best Local Similarity 100.0%; Pred. No. 1.1e-28;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTGAGCTGAGTGTACACCCAGCTTGGCTCCCAAGTCTGGGATTACAG 2109
 DB 163582 TCGAAGCTGAGCTGAGTGTACACCCAGCTTGGCTCCCAAGTCTGGGATTACAG 163641

QY 2110 GTGTGAGCCAC 2120
 |||||

DB 163642 GTGTGAGCCAC 163652

```

RESULT 66
AC019118
LOCUS
DEFINITION Homo sapiens BAC clone RP11-513H7 from 2, complete sequence.
AC019118
VERSION
AC019118.7 GI:18464338
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 188914)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL
MEDLINE
99063792
PUBMED
9847074
REFERENCE
2 (bases 1 to 188914)
Biellik, L., Haakenson, W. and Marquis-Homeyer, C.
The sequence of Homo sapiens BAC clone RP11-513H7
Unpublished (2001)
3 (bases 1 to 188914)
Waterston, R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 188914)
Waterston, R.H.
Direct Submission
Submitted (01-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 188914)
Waterston, R.
Direct Submission
Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 1, 2002 this sequence version replaced gi:14280290.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0513H07

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. Additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-513 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Fenggen, E.,

Tateno, M., Cataneese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-618J22. Actual start of this clone is at base position 1 of RP11-513H7, actual end is at base position 188914 of RP11-513H7.

There is a region from 45729 to 45825 that is covered by a single M13 subclone. There is a single sequence repeat from 151784 to 152156. Polymorphisms have been identified between AC073502 and AC019118. Data from AC073502 was used to finish this clone, AC019118.

FEATURES

source

1. 188914
 Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-513H7"

/clone_1b="RP11-11"

1185..1482

/rpt_family="Alu"

1666..2194

/rpt_family="L1"

2560..2759

/rpt_family="L2"

3056..3241

/rpt_family="L2"

3811..4108

/rpt_family="Alu"

4574..4783

/rpt_family="MER1_type"

6533..6538

/note="match to EST B1460036 (NID:915250692)"

6669..6925

/rpt_family="MALR"

6969..7147

/rpt_family="MALR"

8421..8674

/rpt_family="L1"

8809..9335

/rpt_family="L1"

9127

/note="match to EST AL523930 (NID:912787423)"

9376..9526

/rpt_family="L1"

9527..10469

/rpt_family="ERVK"

10470..10481

/rpt_family="L1"

10668..11063

/rpt_family="MALR"

11056..11718

/note="match to EST AL523930 (NID:912787423)"

11059..11880

/note="match to EST AL523931 (NID:912787424)"

11134..11334

/rpt_family="L1"

11369..11454

/rpt_family="ERVU"

11606..11654

/rpt_family="Mariner"

15104..15279 (TG)n"

/rpt_family="TG)n"

15171..15243

/rpt_family="ERV1"

15318..15650

/rpt_family="TATG)n"

repeat_region

15658..16018

/rpt_family="TG)n"

repeat_region

16687..16733

/rpt_family="(A)n"

repeat_region

16736..17047

/rpt_family="Alu"

repeat_region

17232..17312

/rpt_family="MIR"

repeat_region

17602..18319

/rpt_family="L1"

repeat_region

18416..18635

/rpt_family="MIR"

repeat_region

18642..18948

/rpt_family="L1"

repeat_region

19097..19223

/rpt_family="L1"

repeat_region

19250..19439

/rpt_family="MIR"

repeat_region

19816..20208

/rpt_family="ERVU"

repeat_region

20242..20598

/rpt_family="MALR"

repeat_region

21201..21237

/rpt_family="(TGAG)n"

repeat_region

21313..21421

/rpt_family="L2"

repeat_region

22746..23426

/note="match to EST BF511221 (NID:911594519)"

repeat_region

23189..23565

/rpt_family="L1"

repeat_region

23569..23876

/rpt_family="Alu"

misc_feature

23575..23591

/note="match to EST A1815922 (NID:95431468) au43e07.x1"

misc_feature

23579..23590

/note="match to EST BB858494 (NID:910373581)"

repeat_region

23932..23958

/rpt_family="AT-rich"

repeat_region

24372..24528

/rpt_family="MALR"

repeat_region

24771..25584

/rpt_family="L1"

repeat_region

25592..25772

/rpt_family="MER2_type"

repeat_region

25773..26245

/rpt_family="L1"

repeat_region

26246..26366

/rpt_family="Alu"

repeat_region

26371..26664

/rpt_family="Alu"

repeat_region

26668..27064

/rpt_family="L1"

repeat_region

27075..27205

/rpt_family="L1"

repeat_region

27588..27879

/rpt_family="L1"

repeat_region

27889..28100

/rpt_family="L1"

repeat_region

28169..28203

Query Match

Best Local Similarity 100.0%; Pred. No. 1.1e-28;

Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2.8%; Score 71; DB 9; Length 188914;

2050 TCGAATCTGACCTGAGTATCCACCTTGGCTCCCAAGTGGGATTACAG 2109

DB 160446 TCGAATCTGACCTGAGTATCCACCTTGGCTCCCAAGTGGGATTACAG 160505

2110 GTGTGAGCCAC 2120

DB 160506 GTGTGAGCCAC 160516

```

RESULT 67
AC130472/c 193605 bp DNA linear HTG 10-AUG-2002
LOCUS Homo sapiens chromosome 3 clone RP11-395P16, WORKING DRAFT
DEFINITION AC130472 AC069495
SEQUENCE, 5 unordered pieces.
ACCESSION AC130472.1 GI:22203281
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILL; HTGS_ACTIVEFIN.
KEYWORDS human
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 193605)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE Direct Submision
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193605)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE Direct Submision
JOURNAL Submitted (10-AUG-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Aug 10, 2002 this sequence version replaced gi:8576290.
COMMENT ----- Genome Center
Center: University of Washington Genome Center
Center Code: UMGSC
Web site: http://www.genome.washington.edu
Contact: umgsc@u.washington.edu
Drafting Center: BCM
----- Project Information
Center project name: chr-3
Center clone name: RP11-395P16 (bc0403)
----- Summary Statistics
Sequencing vector: plasmid; 54% of reads
Sequencing vector: unknown; 46% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990119
Consensus quality: 190960 bases at least Q40
Consensus quality: 191966 bases at least Q30
Consensus quality: 192639 bases at least Q20
Insert size: 193205; sum-of-contigs
Quality coverage: 11.1x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4015: contig of 4015 bp in length
* 4016 4115: gap of unknown length
* 4116 18413: contig of 14298 bp in length
* 18414 18513: gap of unknown length
* 18514 33150: contig of 14637 bp in length
* 33151 33250: gap of unknown length
* 33251 65606: contig of 32356 bp in length
* 65607 65706: gap of unknown length
* 65707 193605: contig of 127899 bp in length.
*
* Location/Qualifiers
* 1. 193605
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="3"
* /clone="RP11-395P16"
* /clone_1ib="RPCT human BAC library 11"
* 1. 4015
* /note="assembly_name:Contig23"
* 4116. 18413
* /note="assembly_name:Contig24"
* 18514. 33150
* /note="assembly_name:Contig25"
* 33251. 65606
misc_feature
misc_feature
misc_feature

```

```

misc_feature /note="assembly_name:Contig26"
65707. 193605
BASE COUNT 60749 a 41346 c 40148 g 50958 t 404 others
ORIGIN
Query Match 2.8%; Score 71; DB 2; Length 193605;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2050 TCGAAGCTCTGACCTGAGTATGACCCACCTGGGCTCCCAAGTGTGGGATTACAG 2109
Db 147187 TCGAAGCTCTGACCTGAGTATGACCCACCTGGGCTCCCAAGTGTGGGATTACAG 147128
QY 2110 GTGTGAGCCAC 2120
Db 147127 GTGTGAGCCAC 147117
RESULT 68
AC073641 197709 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-368B14 from 2, complete sequence.
DEFINITION AC073641
ACCESSION AC073641.3 GI:15638827
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 197709)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 197709)
AUTHORS Swearingen, S. and Maupin, R.
TITLE The sequence of Homo sapiens BAC clone RP11-368B14
JOURNAL Unpublished (2001)
JOURNAL 3 (bases 1 to 197709)
REFERENCE 3 (bases 1 to 197709)
AUTHORS Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (27-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 197709)
AUTHORS Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 197709)
AUTHORS Waterston, R.
TITLE Direct Submision
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 18, 2001 this sequence version replaced gi:13431168.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.edu
----- Summary Statistics
Center project name: H_NH0368E14
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.
This sequence was finished as follows unless otherwise noted:

```

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone, and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tateo, M., Caranese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC068343; the clone sequenced to the right is RP11-488C22, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-368E14; actual end is at base position 197709 of RP11-368E14.

Data from AC012460, AC016701, and AC012664 was used to finish this clone, AC073641.

FEATURES

```

source
    1..197709
        /location="Qualifiers"
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="2"
        /map="2"
        /clone="RP11-368E14"
        /clone_1lb="RPCI-11"
        1..69
            /rpt_family="ERV1"
            214..516
                /rpt_family="ERV1"
            721..921
                /rpt_family="Alu"
            774..904
                /note="match to EST AW753869 (NID:97668801)"
            796..919
                /note="match to EST BG958381 (NID:914376552)"
            813..951
                /note="match to EST BG958432 (NID:914376603)"
            925..954
                /rpt_family="Alu"
            1308..1440
                /rpt_family="AT_rich"
            1797..1959
                /rpt_family="MIR"
            1968..2257
                /rpt_family="MIR"
            2253..2273
                /rpt_family="Alu"
            /note="similar to Homo sapiens EST BE972557 (NID:910585893)"
            repeat_region
                2259..2563
                    /rpt_family="Alu"
            repeat_region
                2261..2284
                    /rpt_family="Alu"
            repeat_region
                2661..2682
                    /rpt_family="AT_rich"
            repeat_region
                3021..3046
                    /rpt_family="AT_rich"
            repeat_region
                3127..3184
                    /rpt_family="L2"
            repeat_region
                3185..3609
                    /rpt_family="L2"
            repeat_region
                3610..3658
                    /rpt_family="ERV1"
            repeat_region
                3659..3866
                    /rpt_family="L2"
            repeat_region
                3837..3874
                    /rpt_family="Alu"
            repeat_region
                4235..4520
                    /rpt_family="TAAAA)n"
            repeat_region
                4868..5102
                    /rpt_family="Alu"
            repeat_region
                4983..5393
                    /rpt_family="Alu"
            misc_feature
                /note="match to EST H84027 (NID:91062698) yv88c12.r1"
            repeat_region
                5409..5451
                    /rpt_family="Alu"
            repeat_region
                5452..5623
                    /rpt_family="Alu"
            repeat_region
                5884..6192
                    /rpt_family="Alu"
            repeat_region
                6489..6615
                    /rpt_family="MIR"
            repeat_region
                6743..6837
                    /rpt_family="MIR"
            repeat_region
                7062..7083
                    /rpt_family="MIR"
            repeat_region
                7079..7364
                    /rpt_family="AT_rich"
            repeat_region
                7367..7403
                    /rpt_family="Alu"
            repeat_region
                7376..7570
                    /rpt_family="TTTTA)n"
            repeat_region
                7625..7934
                    /rpt_family="Alu"
            repeat_region
                7937..7988
                    /rpt_family="Alu"
            repeat_region
                7986..8208
                    /rpt_family="MIR"
            repeat_region
                8267..8494
                    /rpt_family="MIR"
            repeat_region
                8981..9035
                    /rpt_family="Alu"
            repeat_region
                9307..9606
                    /rpt_family="AT_rich"
            repeat_region
                9638..9664
                    /rpt_family="Alu"
            repeat_region
                9742..9847
                    /rpt_family="TTTTC)n"
            repeat_region
                9903..9955
                    /rpt_family="Alu"
            repeat_region
                9957..10702
                    /rpt_family="CT_rich"
            repeat_region
                10172..10222
                    /rpt_family="Al"
            repeat_region
                10602..10622
                    /rpt_family="(TG)n"
            repeat_region
                10703..10872
                    /rpt_family="AT_rich"
            repeat_region
                10880..10995
                    /rpt_family="Alu"
            repeat_region
                11080..11205
                    /rpt_family="Alu"
            repeat_region
                11417..11543
                    /rpt_family="MER1_type?"
            repeat_region
                11525..11702
                    /rpt_family="TTTTC)n"
            repeat_region
                12025..12078
                    /rpt_family="Alu"
            repeat_region
                12462..12747
                    /rpt_family="AT_rich"
            repeat_region
                12728..12757
                    /rpt_family="Alu"
            repeat_region
                12823..12912
                    /rpt_family="CAAAA)n"
            repeat_region
                /rpt_family="MIR"

```



```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0323J07
----- Summary Statistics -----
Sequencing vector: M13: +
Sequencing vector: plasmid: +
Chemistry: Dye-primed ET; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 196368 bases at least Q40
Consensus quality: 196955 bases at least Q30
Consensus quality: 197431 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 198448; sum-of-contigs
Quality coverage: 8.49 in Q20 bases; agarose-fp
Quality coverage: 8.15 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 16894: contig of 16894 bp in length
* 16895 16994: gap of unknown length
* 16995 40481: contig of 23487 bp in length
* 40482 40581: gap of unknown length
* 40582 87273: contig of 46692 bp in length
* 87274 87373: gap of unknown length
* 87374 198748: contig of 111375 bp in length.
*
* Location/Qualifiers
*
* source
*   1..198748
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="5"
*   /clone="RP11-323J7"
*   1..16894
*   /note="assembly_name:Contig20"
*   clone_end:SP6
*   vector_side:left
*   16995..40481
*   /note="assembly_name:Contig21"
*   40582..87273
*   /note="assembly_name:Contig22"
*   clone_end:77
*   vector_side:right
*   87374..198748
*   /note="assembly_name:Contig23"
*
* BASE COUNT  61732 a 38112 c 37855 g 60748 t 301 others
*
* ORIGIN
*
* Query Match      2.8%; Score 71; DB 2; Length 198748;
* Best Local Similarity 100.0%; Pred. No. 1,1e-28;
* Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY  2050 TCGAAGCTCTGACCTGAGTATCCACCACCTTGCGCTCCCAAGTCTGGGATTACAG 2109
*      |||||
* Db  71423 TCGAAGCTCTGACCTGAGTATCCACCACCTTGCGCTCCCAAGTCTGGGATTACAG 71364
*
* QY  2110 GTGTGAGCCAC 2120
*      |||||
* Db  71363 GTGTGAGCCAC 71353
*
* RESULT 72
* LOCUS      AP000494      200000 bp      DNA      linear      PRI 28-SEP-1999
* DEFINITION Homo sapiens genomic DNA, chromosome 3p21.3, clone:301 to 308,

```

```

----- anti-oncogene region, section 2/5.
ACCESSION      AP000494
VERSION        AP000494.1 GI:5926671
KEYWORDS
SOURCE
ORGANISM
  Homo sapiens
  Homo sapiens DNA.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases)
  Nakamura,Y., Isomura,M., Daigo,Y., Tamari,M. and Ishikawa,S.
  DNA sequence analysis of a 1.2-Mb region on chromosome 3p21.3
  Published Only in DataBase (1999)
  2 (bases 1 to 200000)
  Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  Direct Submission
  Submitted (21-SEP-1999) Miki Hirakawa, Japan Science and Technology
  Corporation (JST), Advanced Database Department, 5-3, Yonbancho,
  Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mikah@tokyo.jst.go.jp,
  URL: http://www-alls.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
  Fax:81-3-5214-8470)
  This sequence is conducted by Japanese Foundation for Cancer
  Research as a JST sequencing Team.
  Principal Investigator: Yusuke Nakamura Ph.D
  Phone: +81-3-5449-5372, Fax: +81-3-5449-5433,
  yusuke@jgc.ims.u-tokyo.ac.jp
  The sequence is submitted by Human Genome Sequencing in ALIS
  project of JST
  Japan Science and Technology Corporation (JST)
  5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
  For further information about this sequences, please visit our
  sequence archive Web site (http://www-alls.tokyo.jst.go.jp/HGS/top.
  html) or send email to webmaster@www-alls.tokyo.jst.go.jp.
*
* Location/Qualifiers
*
* source
*   1..200000
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /map="3p21.3"
*   /chromosome="3"
*   complement(112666..112810)
*   /note="STS9202; The location is between each flanking site
*   of PCR primers."
*   /db_xref="GDB:456993"
*
* BASE COUNT  52190 a 42809 c 46061 g 58756 t 184 others
*
* ORIGIN
*
* Query Match      2.8%; Score 71; DB 9; Length 200000;
* Best Local Similarity 100.0%; Pred. No. 1,1e-28;
* Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY  2050 TCGAAGCTCTGACCTGAGTATCCACCACCTTGCGCTCCCAAGTCTGGGATTACAG 2109
*      |||||
* Db  139609 TCGAAGCTCTGACCTGAGTATCCACCACCTTGCGCTCCCAAGTCTGGGATTACAG 139668
*
* QY  2110 GTGTGAGCCAC 2120
*      |||||
* Db  139669 GTGTGAGCCAC 139679
*
* RESULT 73
* LOCUS      AC012460/C      200885 bp      DNA      linear      HTG 21-FEB-2001
* DEFINITION Homo sapiens chromosome 2 clone RP11-536G14, WORKING DRAFT
* SEQUENCE 9 unordered pieces.
* ACCESSION      AC012460
* VERSION        AC012460
* KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
* SOURCE
* ORGANISM
*   Homo sapiens
*   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
*   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
*   1 (bases 1 to 200885)
*   Waterston,R.H.

```

TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 200885)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Feb 21, 2001 this sequence version replaced gi:8570287.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0536G14
----- Summary Statistics -----

Sequencing vector: MJ3; 73%
Chemistry: Dye-terminator Big Dye; 27% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195782 bases at least Q40
Consensus quality: 197726 bases at least Q30
Consensus quality: 198568 bases at least Q20
Insert size: 20200; agarose-1p
Insert size: 20085; sum-of-contigs
Quality coverage: 4.91 in Q20 bases; agarose-1p
Quality coverage: 5.28 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1894: contig of 1894 bp in length
* 1895 1894: gap of unknown length
* 1895 8748: contig of 6754 bp in length
* 8749 8848: gap of unknown length
* 8849 14924: contig of 6076 bp in length
* 14925 15024: gap of unknown length
* 15025 26086: contig of 11062 bp in length
* 26087 26186: gap of unknown length
* 26187 38993: contig of 12807 bp in length
* 38994 39093: gap of unknown length
* 39094 51696: contig of 12602 bp in length
* 51696 51795: gap of unknown length
* 51796 79679: contig of 27884 bp in length
* 79680 79779: gap of unknown length
* 79780 112476: contig of 32697 bp in length
* 112477 112576: gap of unknown length
* 112577 200885: contig of 88309 bp in length.

FEATURES

Location/Qualifiers

1..200885
/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/clone="RP11-536G14"

1..1894

/note="assembly_name:Contig20"

1895..8748

/note="assembly_name:Contig21"

8849..14924

/note="assembly_name:Contig22"

clone_end:17

vector_side:right

15025..26086

/note="assembly_name:Contig23"

26187..38993

/note="assembly_name:Contig24"

* misc_feature

/note="assembly_name:Contig24"

clone_end:SP6
vector_side:right
39094..51695
/note="assembly_name:Contig25"
51796..79679
/note="assembly_name:Contig26"
79780..112476
/note="assembly_name:Contig27"
112577..200885
/note="assembly_name:Contig28"
BASE COUNT 56415 a 42233 c 42806 g 58627 t 804 others
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 200885;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAATCTTACTGATTCACCCACCTTGCTCCCAAGTCTGGATTACG 2109

Db 146180 TCGAATCTTACTGATTCACCCACCTTGCTCCCAAGTCTGGATTACG 146121

Qy 2110 GTGTGAGCCAC 2120

Db 146120 GTGTGAGCCAC 146110

RESULT 74
AC005071/c 209382 bp DNA linear PRI 24-MAR-1999

LOCUS Homo sapiens clone RG161A02, complete sequence.
DEFINITION AC005071
ACCESSION AC005071.2 GI:4508112
VERSION
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 209382)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 209382)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 209382)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Mar 24, 1999 this sequence version replaced gi:3212920.
FEATURES
Location/Qualifiers
1..209382
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 53274 a 51565 c 52404 g 52139 t
ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 209382;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAATCTTACTGATTCACCCACCTTGCTCCCAAGTCTGGATTACG 2109

Db 193758 TCGAATCTTACTGATTCACCCACCTTGCTCCCAAGTCTGGATTACG 193699

Qy 2110 GTGTGAGCCAC 2120

Db 193698 GTGTGAGCCAC 193688

RESULT 75
AC006238
LOCUS AC006238 211945 bp DNA linear PRI 31-JAN-1999
DEFINITION Homo sapiens chromosome 18, clone hRPK.474_N_24, complete sequence.
AC006238
VERSION AC006238.1 GI:4204704
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone hRPK.474_N_24
JOURNAL Unpublished
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donegan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatae,A., Lehoczy,J., Macdonald,P., Margis,N., McEwan,P., McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Nilotf,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donegan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heathford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatae,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Margis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Nilotf,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>.
FEATURES
source
1.211945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="hRPK.474_N_24"
/clone_lib="RPCI-11 human BAC library"
complement(32..183)
repeat_region
/rpt_family="LIM3C"
2108..2211
/rpt_family="TG)n"
repeat_region
/rpt_family="CA)n"
3040..3092
complement(6746..7851)
repeat_region

repeat_region
/rpt_family="L2"
9008..9186
/rpt_family="MIR"
repeat_region
/rpt_family="ALUc"
9608..9902
/rpt_family="ALUc"
9903..9931
/rpt_family="AT_rich"
9946..9992
/rpt_family="Alu"
repeat_region
/rpt_family="TA)n"
12001..12169
complement(12218..12291)
/rpt_family="MIR"
repeat_region
complement(13995..14112)
/rpt_family="LIM4"
complement(14113..14412)
/rpt_family="ALUsg1"
complement(14413..14566)
/rpt_family="LIM4"
repeat_region
/rpt_family="MER58A"
14608..14822
complement(14831..15211)
/rpt_family="MLT1"
15540..15705
/rpt_family="MER5B"
repeat_region
/rpt_family="LIPAG"
15836..16582
complement(16583..17089)
/rpt_family="MER9"
repeat_region
/rpt_family="LIPAG"
17090..22509
/rpt_family="LIPAG"
22777..23406
/rpt_family="LIMC"
repeat_region
complement(23983..24178)
/rpt_family="MLT1"
repeat_region
complement(25899..26406)
/rpt_family="MLT1D"
26790..26810
/rpt_family="CCTG)n"
complement(30033..31005)
/rpt_family="L2"
repeat_region
/rpt_family="MLT1D"
31346..31959
/rpt_family="MLT1D"
31960..32236
/rpt_family="MER44D"
complement(32237..32591)
/rpt_family="MER47A"
32592..32991
/rpt_family="MER44D"
complement(34134..34349)
/rpt_family="MIR"
34424..34847
/rpt_family="LIMC4"
complement(34871..34969)
/rpt_family="ALUc/PRAM"
35021..35144
/rpt_family="TRATATG)n"
35256..35356
/rpt_family="AT_rich"
35621..35658
/rpt_family="AT_rich"
complement(35783..35897)
/rpt_family="MIR"
complement(36089..36316)
/rpt_family="MIR"
36787..37087
/rpt_family="ALUy"
37383..37556
/rpt_family="MER102"
37578..37607
/rpt_family="CA)n"
39356..39383
/rpt_family="(T)n"

Job time : 17913 secs

```

repeat_region      39936..40082
/rpc_family="MIR"
repeat_region      40495..40637
/rpc_family="MER5B"
repeat_region      41783..42073
/rpc_family="AluSg"
repeat_region      42856..42965
/rpc_family="MIR"
repeat_region      45120..45328
/rpc_family="L1MB8"
repeat_region      46246..46398
/rpc_family="MER5B"
repeat_region      complement(46520..46819)
/rpc_family="AluSg"
repeat_region      complement(47951..48228)
/rpc_family="MER33"
repeat_region      50241..50533
/rpc_family="AluO"
repeat_region      51158..51215
/rpc_family="TA)n"
repeat_region      51624..52514
/rpc_family="Tiggr3b"
repeat_region      52515..52741
/rpc_family="MER30"
repeat_region      52742..52934
/rpc_family="Tiggr3b"
repeat_region      53342..53375
/rpc_family="AT_rich"
repeat_region      53450..53754
/rpc_family="AluX"
repeat_region      54957..55256
/rpc_family="AluYB8"
repeat_region      55888..55921
/rpc_family="AT_rich"
repeat_region      56168..56581
/rpc_family="WSTA"
repeat_region      56833..56920
/rpc_family="L1MA9"
repeat_region      57982..58106
/rpc_family="MER103"
repeat_region      58703..59203
/rpc_family="L1MC5"
repeat_region      complement(59212..59426)
/rpc_family="MER20"
repeat_region      59443..59650
/rpc_family="L1MC5"
repeat_region      59651..59937
/rpc_family="AluSp"
repeat_region      59938..60333
/rpc_family="L1MC5"
repeat_region      61866..61886
/rpc_family="AT_rich"
repeat_region      complement(62370..62523)
/rpc_family="AluO"
repeat_region      62524..62557
/rpc_family="(TTTG)n"
repeat_region      complement(62558..62676)
/rpc_family="AluO"

```

```

Query Match      2.8%; Score 71; DB 9; Length 211945;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAATCCTGAGCTCAGTATCCACCACTGCTCCCAAGTGGATTACAG 2109
      |||
Db 109238 TCGAATCCTGAGCTCAGTATCCACCACTGCTCCCAAGTGGATTACAG 109297

QY 2110 GTGTGAGCCAC 2120
      |||
Db 109298 GTGTGAGCCAC 109308

```

Search completed: March 30, 2003, 21:28:58